

Science and more for TB

Traditional systems for developing drugs are failing spectacularly to deliver the goods in the fight against tuberculosis. Innovative public-private collaborations point the way forward.

The diseases that afflict the developing world defy the imagination in their scale. One third of the world's population — 2 billion people — are infected with *Mycobacterium tuberculosis*, and 16 million have active tuberculosis (TB). Shockingly, TB hit an all-time high in 1999 with 8 million new cases — 95 per cent of them in developing countries — and 2 million deaths. The disease is spreading rapidly through Russia. The toll is set to rise; AIDS activates the dormant form of the disease, while multidrug resistance is spreading across the planet. Figures for malaria are similar, and AIDS is raging out of control throughout the developing world.

If global health is left to market forces, historians will remember this era as one in which humanity stood idly by while half the planet languished in sickness. Fortunately some enlightened individuals have realized this, and are driving forward new models for drug and vaccine discovery. The latest sign of this trend is the creation, at an international meeting last week in Cape Town, South Africa, of a Global Alliance for TB Drug Development, by a consortium of scientists, donor agencies and pharmaceutical companies (see page 692).

To be of practical use, any drug or vaccine must be cheap to produce and administer. Institutional innovation to overcome these obstacles is underway: as well as the public sector underwriting research to compensate for market failure, the less tangible goal of a tight coupling of scientists, regulators, donors and companies around shared objectives is also taking place. The alliance will need all the determination it can muster, and luck too.

Science alone is not enough. By coordinating research and taking into account the needs of those in the field, the alliance has a good chance of achieving its objective of a new TB drug by 2007. But also, as one delegate in Cape Town put it: "New drugs without effective control programmes = new resistance."

The time is right. The genome of *M. tuberculosis* is publicly available (http://www.sanger.ac.uk/Projects/M_tuberculosis), and a raft of technologies, such as DNA chips to identify genes involved in infection, are under development. Just as important as treatments for full-blown TB are drugs and diagnostics for the huge reservoir of carriers of the dormant form. Looking for genes activated during disease activation may help identify drug targets. Understanding how the dormant phase escapes drugs and the immune system is also a challenge.

Imaginative use of new technologies will be vital, because the venture's success hangs on cutting the costs of developing a new drug — typically US\$300 to \$500 million — to \$100 million. Big drug companies will probably subsidize the venture by making their vast combinatorial-chemistry libraries and high-throughput screening technologies freely available. They will also be needed for getting bulk quantities of drugs distributed far and wide.

But the alliance may be going too far in defining its programme so as to win over the participation of the large drug companies. The latter's bureaucracies are ill-suited to rapid change and fast decision-making, and could defeat the alliance's ambition of being a lightweight structure. For innovation, flexibility and cost-cutting, the alliance would do better to shift its focus towards small pharmaceutical and biotechnology companies.

Drug companies need to make their drugs more affordable in developing countries. But developing countries must also play their part. The government of relatively rich South Africa is already equivocating over the known value of anti-AIDS drugs when one in four of its citizens is HIV-positive (see page 692). Even acknowledging the political difficulties, South Africa and other developing countries need reminding that the cost of business as usual will be much more than that of investing in the development of drugs and vaccines. ■

Fossil smuggling unopposed

Trading at fossil shows requires law enforcement.

A paleontology débâcle unfolding in the United States over a Chinese bird fossil from the early Cretaceous will probably go down as the classic example of the danger to science of amateurs dealing in smuggled specimens (see pages 689–690). The tortured tale of the fossil would be funny were it not for the fact that a potentially valuable slice of history may have been damaged.

The specimen was sold a year ago at a mineral and fossil show in Tucson, Arizona, to supporters of a small museum in Utah. Subsequently, the operator of the museum — who has no formal palaeontology training and little experience with peer-reviewed journals — acted as the lead author of an article submitted to *Nature* and *Science*.

Despite the help of some well-meaning palaeontologists, the publishing attempt failed. Next, *National Geographic* magazine embarrassed itself by naively and hastily publishing an article — described as "sensationalistic, unsubstantiated tabloid journalism" by a leading paleontologist — sprinkled with dubious assertions.

But the real culprits are the fossil shows in the United States and Europe that provide a market for smuggled national treasures. Operators of the US shows say they try to make sure that dealers do not sell smuggled or illegal goods. The dealers have documents for the specimens that purport to permit their sale as scientific exchanges. But does anyone truly believe that someone can buy a transferable right to a scientific exchange in a cheap hotel room?

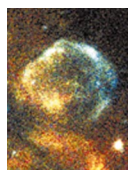
US federal authorities say that there are laws prohibiting trade in fossils smuggled from other nations, such as China. Scandalously, the law goes unenforced. At the Tucson show, state and city police were serving as security, even pawing through boxes of ivory from dubious sources for a specimen to take home. Hardly reassuring.

International agreements are needed to eliminate illicit trade in fossils, and scientists and museums should push for such pacts. But the power is already there to protect palaeontology from the illegal deals that have plagued the field for decades. It should be applied. ■





Tuberculosis
Public-private consortium aims to discover new drugs
p692



X-ray astronomy
Europe's Newton observatory firing on all cylinders
p693



McCain
Defends decision to change tack on fetal-tissue research
p694



Oxford
Enquiries broaden into affairs of zoology professor
p695

Feathers fly over Chinese fossil bird's legality and authenticity

San Diego

The US palaeontology community has been rocked by a Chinese 'bird' fossil that may be a new species, but that many suspect to be a composite of more than one fossil that was smuggled illegally out of China.

The fossil was bought last year at an Arizona mineral show for \$80,000 on behalf of a small Utah museum. It has since been valued for insurance purposes at \$1.6 million, and has been heralded as an important link in dinosaur and bird evolution — in particular, in an article in *National Geographic* last autumn.

But Chinese scientists want the Utah museum to repatriate it immediately. And, in a move which has left the magazine deeply embarrassed, some palaeontologists are also saying that the fossil's tail may come from another specimen.

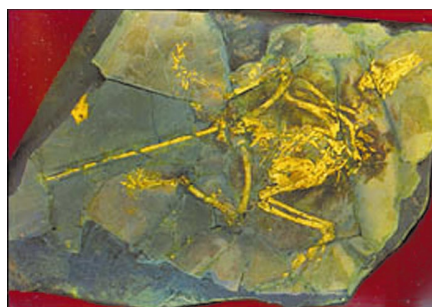
"This is a disaster for science," says Xu Xing, a graduate student at the Institute of Vertebrate Paleontology and Paleoanthropology in Beijing, who collaborated on a paper describing the fossil and has been involved in repatriation negotiations.

The fossil is believed to have come from the Liaoning area in northeast China, where many bird-like specimens dating from the Cretaceous period have been discovered.

An article in the November 1999 issue of *National Geographic* claims that the specimen is one of the "missing links in dinosaur evolution". But computerized tomography tests have shown that the fossil's tail may have been added in a bid to increase its black-market value. The magazine is to publish a note in its March issue stating that the specimen is "a composite", and that tests have "revealed anomalies in the fossil's construction".

In the November article, the magazine gave the specimen a name — *Archaeoraptor liaoningensis* — even though it had not been formally described in a peer-reviewed journal. This angered scientists such as Storrs Olson, curator of birds at the US National Museum of Natural History in Washington DC, who has described the move as "the worst nightmare of many zoologists".

Magazine officials say that the publication of the name and details of the fossil was a



Dead duck? 'Archaeoraptor' may be the tail of a primitive bird stuck to the body of a dinosaur.

"mistake" that resulted from marketing efforts leapfrogging an unsuccessful bid for scientific publication — one of several mis-cues during the fossil's journey through America.

Bill Allen, *National Geographic's* editor, said in a letter to a scientist involved that he "was totally outraged" by what occurred, acknowledging that it has damaged the magazine's credibility. Chris Sloan, who wrote the fossil story, says that the magazine has now agreed not to publish the name of a specimen until a scientific journal has described it.

Exactly how the fossil left China is unknown, although Xu and other scientists insist it must have been smuggled out. Its trail in America begins in the autumn of 1998, when scientists gathered in a Utah ski resort for the annual meeting of the Society

of Vertebrate Paleontology heard rumours of an exciting bird fossil available on the private market.

The fossil — which scientists now believe was the one featured in *National Geographic* — was offered for sale in February 1999 at the annual Gem, Mineral and Fossil Showcase of Tucson, Arizona, reportedly the world's largest gathering of fossil dealers.

The purchase of the fossil was organized by Stephen Czerkas, a self-educated dinosaur enthusiast and artist who runs the non-profit Dinosaur Museum in the small town of Blanding, Utah.

Czerkas convinced patrons of the Dinosaur Museum, including trustee Dale Slade, to put up \$80,000 to buy the specimen. Slade, a businessman who provides material for semiconductors, says that documents exist showing that the fossil is part of a museum exchange programme and was not smuggled; but he declined to make such records available to *Nature*.

Czerkas says that he bought the fossil to save it for science, planning to return it to China after display. Sloan says the fossil was to become "the crown jewel" in the Blanding museum, calling the facility "the brainchild" of Czerkas and his artist wife, Sylvia.

Sloan says that last spring, while working on his *National Geographic* article, he went to Utah and convinced Czerkas to repatriate the fossil, telling him the magazine would

UK scientists under pressure to please

London

One in three UK scientists working in government or recently privatized laboratories has been asked to alter research findings, according to a survey by a trade union that represents government-employed researchers.

Thirty per cent of respondents had been asked to modify their conclusions or advice. Reasons given were: to suit the customer's preferred outcome (17 per cent); to obtain further contracts (10 per cent); and to

prevent publication (3 per cent).

The Institute of Professionals, Managers and Specialists (IPMS), which represents scientists in government departments, research councils and private companies, surveyed over 500 respondents from a range of employers.

An IPMS spokesperson said that the findings were "clearly significant at a time when the independence of scientists, both in and out of government, is increasingly being questioned".

Natasha Loder

not write about it unless he did so. Czerkas agreed to this, says Sloan.

Meanwhile, Czerkas was working on a paper about the fossil that he hoped to publish in a major journal. Last summer, the plan was to publish the manuscript naming the specimen, with the *National Geographic* article following on its heels.

Although he does not hold a university degree, Czerkas says that he is widely respected in the scientific community for helping major museums create dinosaur displays. These contacts led him to his collaborators: Philip Currie, a palaeontologist at the Royal Tyrrell Museum of Palaeontology in Canada who has close links to China; Xu, who Currie recruited to the project; and palaeontologist Timothy Rowe, who has a laboratory at the University of Texas that scans specimens.

According to Rowe, by early August his laboratory had found problems with the fossil — in particular the added tail and questions relating to reconstructed leg bones. Rowe described the problems to Czerkas and Currie at a meeting in his laboratory.

Currie was in Argentina and unavailable for comment last week, but has been quoted in newspaper articles as saying that the whole affair is deeply embarrassing.

Czerkas finally agreed to modify the manuscript to address concerns, says Rowe, and the paper was sent to *Nature*, where it was rejected. With *National Geographic*'s September printing deadline approaching, Czerkas sent the manuscript to *Science*, which sent it out for peer review. Two reviewers rejected it in the first week of September.

Suggesting the possibility that "the specimen was smuggled out of China and illegally purchased", one reviewer suggested that the specimen had been "doctored" by its unknown collectors "to enhance its value". He or she also commented that the specimen should be returned to China before being considered for publication.

Czerkas, the lead author, never fully informed *National Geographic* about the details of those rejections, says Sloan. *National Geographic* learnt of *Science*'s rejection shortly after its deadline for the November issue.

In October, *National Geographic* held a news conference to highlight its November issue, unveil the fossil and announce that its repatriation to China had been agreed.

But according to Xu, Czerkas wanted the fossil to stay in the Utah museum for up to five years before sending it back. He says that Czerkas also sought potentially valuable casts of other Chinese dinosaurs and scientific exchanges in return for the fossil.

Xu says that he has since been negotiating with Czerkas to get the fossil returned immediately, as his superiors desire. He and Sloan say that last month, after newspaper reports had questioned the fossil's authenticity, Czerkas dropped the display demands.

Rowe withdrew from the project in early January, upset that Czerkas hadn't told *National Geographic* earlier about the fossil's composite nature and other issues.

Last week, Slade flew to Washington, picked up the fossil at the *National Geographic* and returned it to Utah. He says that the fossil is undergoing further studies and that an attempt will be made to publish the findings in another journal. He added that the patrons — who he declined to identify —

have yet to transfer ownership of the fossil to the Blanding museum.

Xu may bring a similar fossil to Washington for independent comparison with Czerkas's specimen. Czerkas may then report the findings to a palaeontology meeting in March at the Graves Museum in Florida.

Czerkas says that he will return the fossil to China this spring, possibly at the June meeting of the International Society of Avian Paleontology in Beijing.

Rex Dalton

The biggest, wildest fossil market in the west

Tucson, Arizona

The Gem, Mineral and Fossil Showcase of Tucson, a city-wide street fair of international artefacts, is a major marketplace for fossils, many of which, some claim, have been smuggled out of China and other countries.

Chinese bird fossils from the early Cretaceous that have yet to be described, ancient skulls of rhino and ivory-tusked elephant, and dinosaur eggs were all on sale at the annual two-week event that ended last week.

Some of the dealers selling Chinese fossils say they have documents showing they were legally exported. But US and Chinese scientists say the papers are irrelevant, as the specimens are Chinese cultural treasures.

US customs officials say that the importation, possession and sale of a smuggled specimen may violate various federal laws. But an in-depth legal analysis of each case is needed to determine if a criminal offense was committed, says Customs Special Agent Lisa Fairchild in Washington DC.

About 3,000 dealers — mostly selling gems and minerals — sell their wares in hotel rooms, on patches of ground and in huge tents. Dinosaur eggs and rhino skulls could be found among some of the 24 individually operated shows in Tucson, but this year most of the questionable Chinese fossils were being



Happy hunting ground: A fossil elephant skull, for sale in Tucson.

sold at a show held in a converted hotel.

Chinese fossils large and small were displayed throughout the hotel. One room contained boxes of dozens of dinosaur eggs stacked like eggs in a supermarket.

It was in this setting that that dinosaur enthusiast Stephen Czerkas organized the purchase last year of the controversial bird fossil that found its way into the *National Geographic* magazine last November (see main story).

Czerkas refuses to say who sold the fossil for \$80,000. He says that, after he had found it, a patron went to Tucson to buy the fossil for the small Dinosaur Museum in Blanding, Utah. But the museum's main patron, Dale Slade, says that Czerkas went to Tucson and bought the fossil himself.

Timothy Rowe, a University of Texas palaeontologist who has analysed the fossil, visited Tucson this year to investigate its source. While there he met dealers such as Zhouping Guo, a water-company geologist who

runs the company Sin-Am Bridge out of his home near San Diego, California.

Rowe says that Zhouping knows Czerkas and that during Rowe's visit he produced three Chinese bird fossils that appeared to be species not yet described in the scientific literature.

Zhouping denies selling Czerkas the fossil. He says that he has documents from a Chinese museum proving his specimens were legally exported for scientific exchange, although he refused *Nature*'s request to inspect the documents.

Martin Zinn, who ran the show that included independent firms like Sin-Am Bridge, says he prohibits dealers from selling smuggled or illegal goods, that Sin-Am insisted its specimens were properly documented, and that all was "normal".

A spokeswoman for the Metropolitan Tucson Convention and Visitors Bureau, which organizes the showcase, says the organization knew nothing of illegal sales and was not responsible for independent dealers.

R. D.

Protein structure groups seek to draft common ground rules

Washington

Scientists from Britain, the United States, France, Germany, Israel and Japan are taking preliminary steps towards establishing common procedural guidelines for research in structural genomics. The move could lay the groundwork for a single international structural-genomics effort comparable to the Human Genome Project (HGP).

A meeting to be held in April at the Sanger Centre outside Cambridge, England, will discuss ground rules for collaboration, such as common policies on data release and intellectual property—issues at the heart of collaboration on raw sequencing under the HGP.

The goal of structural genomics is to catalogue the folding properties of all natural proteins. Genes with sequence similarities encode protein structures that resemble one another, and these can be used to group protein shapes into families, reducing the time and expense needed to obtain their structure.

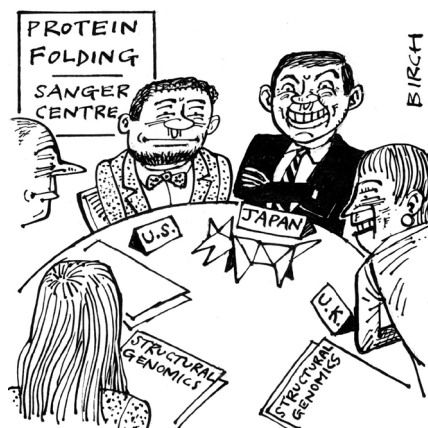
Assembling and analysing an entire set of proteins poses greater technical challenges than sequencing (see *Nature* 402, 703; 1999), but dividing up responsibility among laboratories could cut the project down to a manageable size in each participating country.

The genome project is a “possible model of collaboration” for an international structural-genomics effort, says Michael Morgan, chief executive of the Wellcome Trust’s genome campus.

Barbara Skene, a scientific programme manager at the Wellcome Trust, hopes the meeting will at least result in a database telling scientists which proteins other laboratories are working on, and how far they have got. This could prevent duplication of effort, she says, and help to communicate how different groups solve common problems.

Participants in an international effort will need to agree on standards for the quality of data released into a central database, as well as a time frame—issues heatedly debated during the early stages of the human genome programme. But additional agreement on scientific matters—such as which are the most representative proteins in a family, and how best to divide them between research groups—also need to be considered.

Furthermore, sequencing the genome of a particular organism is highly automated and has a definite end-point, whereas structural genomics is more time-consuming, has a less clear finishing line, and has more potential technical bottlenecks, such as the purification phase, says Skene. Such limitations may even affect the starting point, she adds. For example, membrane proteins, which make up an estimated 25 per cent of known pro-



teins, cannot easily be crystallized, so their structure cannot be determined.

Identifying and addressing procedural similarities between sequencing and structural genomics seems like a good starting point for collaboration, says Marvin Cassman, director of the US National Institute of General Medical Sciences (NIGMS).

NIGMS will launch its own structural-genomics initiative next autumn (see *Nature* 400, 494; 1999). Cassman hopes that the April meeting, which is co-sponsored by NIGMS, will help set international guidelines for data release and sharing of materials.

Cassman admits that reaching consensus on scientific issues, which are not on the agenda for the April meeting, will be tougher. After an initial five-year pilot stage, the NIGMS programme aims to generate 10,000 structures from as many protein families as possible. But some groups intend to characterize every protein structure within small organisms such as the bacterium *Mycoplasma genitalium*, while others may try to obtain as many structures as possible involved in a single cellular process, such as cell division.

These differing approaches could lead to duplication of effort both nationally and internationally. “Some overlap will be good” because it will allow for quality control, says Stephen Burley, a professor at Rockefeller University and investigator with the Howard Hughes Medical Institute. But too much duplication would be wasteful, he adds.

Helen Berman, however, director of the Protein Data Bank at Rutgers University, says organizing an international effort for protein structures “probably won’t be any more or less difficult” than for sequencing efforts.

The need for greater coordination is being watched by the newly renamed Global Science Forum of the Paris-based Organization for Economic Cooperation and Development (OECD).

Paul Smaglik

Austrian body pleads for normal contacts despite EU freeze

Munich

The Austrian Science Fund (FWF), the country’s main research-funding agency, has called on the international scientific community not to turn its back on the country in reaction to the inclusion of the right-wing Freiheitliche Partei Österreichs (FPÖ) in its new coalition government.

The move follows a decision by the European Union to ask its member states to freeze bilateral contacts with Austria. French research minister Claude Allègre said in a television interview last Sunday that all scientific projects between France and Austria were being “frozen”.

“The situation is very worrying,” wrote Arnold Schmidt, president of the FWF, in an open letter to scientists across Europe. “I appeal to all scientists to maintain or increase contact and cooperation with scientists working in Austria.”

The FWF is particularly concerned that researchers will choose to stay away from scientific conferences held in Austria. “If you stay away you will hit science, but not the FPÖ,” he says.

Some scientific societies, such as the European Group of Blood and Marrow Transplantation, are already considering whether to relocate meetings planned for Austria. The Edinburgh International Conference Centre, for example, claims to have been approached by “a number of Austrian organizers” who are considering relocating their events.

Meinhard Regler, vice-president of the Institute for High Energy Physics in Vienna, who is organizing the biannual European Particle Accelerator Conference, says the number of US registrants for this year’s meeting in Vienna is likely to be much lower than usual, although this may be partly due to domestic budget restrictions.

Many Austrian scientists are keen to distance themselves from the FPÖ and its political ideology. “We are totally opposed to the nationalist and xenophobic sentiments expressed by some FPÖ politicians,” said Peter Michor and Klaus Schmidt, scientific directors of the Vienna-based Erwin Schrödinger Institute for Theoretical Physics.

But Kim Nasmyth, the director of the Institute for Molecular Pathology in Vienna—which hosts young researchers from 23 countries—is optimistic that “colleagues everywhere... will not put us in the same boat as some politicians”. Quirin Schiermeier and Patrick Weydt

South African government rejects AZT advice

Cape Town

The South African government announced last week that it has rejected two reports commissioned from its Medicines Control Council (MCC) on the safety of the anti-retroviral drug AZT (zidovudine) after public statements by political leaders on its potential hazards.

Both reports are believed to endorse the use of AZT. But health minister Manto Tshabalala-Msimang told a press briefing held after last week's opening of parliament that the government is not currently prepared to release their contents. She also refused to reveal the findings of a third report compiled by the council, which she said she had only received on 31 January.

The reports were commissioned last October after president Thabo Mbeki had defended the government's decision not to make AZT freely available in the public health system (see *Nature* 402, 3; 1999). His argument that there was a large volume of scientific literature indicating that the drug's toxicity is a danger to health led to an outcry from AIDS activists and researchers in both South Africa and abroad.

Helen Rees, director of the reproductive health research unit at Johannesburg's Baragwanath Hospital and chair of the MCC, said in December that the first report — which she described at the time as “fairly superficial” — found that the benefits of AZT outweighed its risks, and that use of the drug was justified. Rees has not been available for comment this week.

But Tshabalala-Msimang said that the first two reports “were not to our satisfaction”. In particular, she said they had not addressed the risk–benefit assessments of using anti-retroviral drugs in the way requested by her department. She has now asked the MCC “to make available to us information that would assist in determining the risk–benefit assessments”.

The health minister added that she had read only the first few pages of the third report, and was therefore “not quite ready” to comment on it.

But she hoped that this report had addressed the risk–benefit issues in the way that the others had not. Asked if she would make the report public, she replied that legislation made provision for such documents to be made public “when we are finished working on them, if they are required to be published”.

Tshabalala-Msimang is also believed to have received two further reports, one commissioned by the government from the Medicines Research Council, and one from the World Health Organization. Both are understood to recommend the use of AZT in reducing mother-to-child transmission of



Tshabalala-Msimang: wants more details.

HIV/AIDS — a procedure that the government has so far refused to sanction.

The health minister said that the National AIDS Council, which met for the first time last week, has agreed to set up five technical task teams whose briefs would match the priorities of the national AIDS strategic plan adopted in January: prevention; care, treatment and support; research; human rights and legal issues; and social mobilization.

In a later speech to parliament, Tshabalala-Msimang said she and provincial health ministers had accepted guidelines for treat-

ment, throughout the public health system, of “opportunistic” infections that affect HIV/AIDS patients.

But the government's efforts remain under heavy fire from AIDS activists. Mark Heywood, a member of the executive of the AIDS consortium, the country's largest coalition of non-governmental AIDS organizations, criticized both the health minister and the president for failing to meet their legal and moral responsibilities to improve access to health care in South Africa.

“They are wilfully ignoring the best scientific advice, internationally and locally, on the safety and benefits of providing AZT through the public health sector,” he says. “The immediate cost of this impasse is measured in thousands of unnecessary infant HIV infections for which they now carry a very direct responsibility.”

Michael Cherry

Consortium aims to kick-start TB research

Cape Town

A loose consortium of leading scientists, donor agencies and pharmaceutical companies agreed last week to create a joint venture between the public and private sectors to relaunch research into drugs that are able to combat tuberculosis (TB).

Meeting for the first time in Cape Town, South Africa, the Global Alliance for TB Drug Development agreed to draft a scientific blueprint by June of this year. This would aim to reach a consensus on research priorities among participating institutions, and coordinate funding for them.

One-third of the world's population — some 2 billion people — are infected with *Mycobacterium tuberculosis*, the bacterium that causes the disease, and 16 million people have active TB. AIDS activates the dormant form of the disease, and the proportion of carriers who develop active TB is increasing — there are 8 million new cases, and 2 million deaths, each year.

Multidrug resistance is also spreading rapidly throughout the world, raising the spectre of untreatable TB. The last new drug for TB was introduced over 30 years ago, and industry has been reluctant to invest in discovering new families of drugs because of

the financial risks of investing in products destined mainly for developing countries.

The meeting was convened by the Rockefeller Foundation, and co-sponsored by the US National Institutes of Health, the Bill and Melinda Gates Foundation, the Wellcome Trust, the World Health Organization (WHO), the World Bank, the US Centre for Diseases Control and several other donor and research agencies and non-governmental organizations.

The proposed alliance, which is similar to the Multilateral Initiative on Malaria (see *Nature* 388, 219 & 390, 209; 1997), would be coordinated by the WHO.

Several scientists at the meeting pointed out the apparent irony that recognizing the need for new tools has in the past been inhibited by the WHO, which had argued strongly that TB could be contained by the WHO's DOTS strategy (Directly Observed Therapy, Short Course), a cocktail of antibiotics administered daily under medical supervision.

The impracticability of administering this labour-intensive procedure in developing countries and the emergence of multidrug resistance have prompted a U-turn at the WHO. In a video address to the meeting, Gro Harlem Brundtland, WHO director-

general, said that, although DOTS had helped, it was no longer enough and “it is high time to find new and more effective drugs”. Ideally, a drug should have a shorter course and require less supervision by health workers, she said.

As a first step towards launching a common drug-discovery programme for TB between the public and private sectors, akin to the New Medicines for Malaria Initiative, the alliance will prepare a ‘pharmaco-economics’ report by October that will assess in detail the size of the market, current investment, and gaps in corporate research and development.

Giorgio Roscigno, medical director of Hoechst Marion Roussel, told the meeting that one reason that companies have not invested in drugs to combat tuberculosis is that they have underestimated the market.

The alliance plans to use the report as the basis for a business plan for a joint drug-discovery programme that would be announced in December, with the alliance being formally created in early 2001. The programme has the ambitious goal of registering a new drug for TB by 2007 and a second by 2012.

Declan Butler

Europe's X-ray observatory defies the jinx...

Munich

The jinx that destroyed or limited a string of X-ray-astronomy missions launched in the past nine months does not seem to have affected the European Space Agency's (ESA's) new XMM-Newton X-ray observatory. The first images, presented last week, show that the observatory's three X-ray cameras, its optical monitor and its two spectroscopes are operating well.

ESA also announced that it had added the name Newton to the observatory's working name of XMM (X-ray Multi-Mirror), in honour of Isaac Newton, who created the theory of gravity and discovered the Sun's spectrum.

X-ray astronomers are pleased that Newton, which was launched two months ago, has avoided the damage suffered by a similar mission, the US space agency NASA's Chandra. Low-energy protons partly damaged Chandra's cameras a few weeks after launch last July (see *Nature* **401**, 415; 1999).

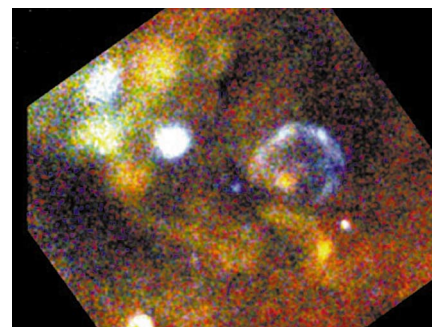
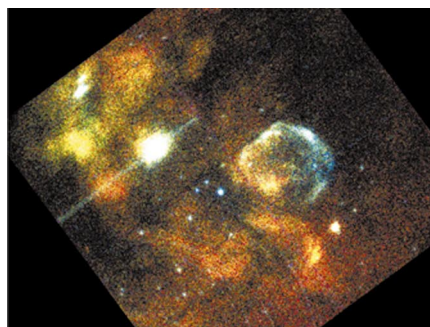
The Newton observatory comprises three telescopes, each consisting of a nest of 58 mirrors, which are designed to detect X-ray photons from deep space. X-rays are emitted by very hot matter, so Newton can observe very hot and violent events in the cosmos, allowing studies of the evolution and death of stars, the structure and nature of the interstellar medium, and the physics of galaxies and quasars.

In addition to the damage to Chandra, X-ray astronomers had previously witnessed the loss of Abrixis, a small German-built X-ray satellite, immediately after its launch last May (see *Nature* **399**, 93; 1999). Last week also saw the loss of Japan's major X-ray mission Astro-E, when its launcher failed to push it into the correct orbit (see right).

Newton scientists say they could have been affected by the problem that struck Chandra, because, like their US colleagues, they had not predicted the damaging effects of one type of cosmic radiation in the deep-space orbit chosen to avoid the influence of the Earth's atmosphere and used by both Chandra and Newton.

Damaging low-energy protons in the Van Allen belts through which both vehicles fly were funnelled by the telescope mirrors onto the cameras. Eight of the ten sensors in Chandra's cameras are 'front-illuminated', making them vulnerable to damage. "Normally these protons are stopped, but obviously they are reflected off the mirrors and we did not anticipate this," says Richard Mushotzky, an astrophysicist at the Goddard Space Flight Center near Washington, and a member of Chandra's science working group.

The damage to Chandra's cameras does not affect the image resolution. But the spectral resolution, which allows the chemical composition of the observed objects to be



In focus: the picture on the left shows one of the first images taken by XMM-Newton, the Tarantula Nebula, part of the Large Magellanic Cloud. The one on the right shows the same region taken previously by ROSAT, a German X-ray satellite that went out of operation over a year ago.

analysed, has been reduced by up to a factor of five. In addition, the mission's ability to analyse X-ray sources that are greater over a field of view than one-fifth of the size of the full moon is also compromised, and these can now be observed, less efficiently, only with the undamaged 'back-illuminated' sensors.

Newton scientists say they have learnt from Chandra's misfortune. Having been more cautious with the original design of their EPIC X-ray cameras, which also include damage-sensitive front-illuminated sensors, they had incorporated a moving shield to allow the cameras to be shut off in the event of any unpredicted radiation damage while passing through the Van Allen belts or during a sunstorm, despite the extra resources this required.

Newton will make up for Chandra's partial loss, as the three big missions were originally complementary, so the weaknesses of one mission could be made up for by strengths in the others. Newton was optimized for sensitivity and for medium-to-high-resolution spectroscopy, Chandra for angular resolution — which has not been affected — and Astro-E for very-high-resolution spectroscopy.

"We are lucky that Newton is well suited to looking at large diffuse sources such as clusters of galaxies, just the capability we lost," says Claude Canizares, an astrophysicist from the Massachusetts Institute of Technology and principal investigator of a Chandra instrument. "But we have completely lost the important piece of the scientific pie that Astro-E would have offered."

Alison Abbott

...but Astro-E is yet another Japanese failure

Tokyo

X-ray astronomers in Japan received a major setback last week with the loss of Astro-E, a joint US-Japanese satellite carrying several innovative X-ray sensors. The failure is the latest in a series of accidents, and could have serious repercussions for Japan's space policy.

The satellite missed its orbit after an engine failure in the first stage of its launch vehicle, an M-V rocket built at Japan's second space agency, the Institute of Space and Astronautical Science (ISAS).

In addition to four X-ray imaging spectrometers, Astro-E carried a novel X-ray spectrometer — the fruit of collaboration between the US space agency NASA's Goddard

Space Flight Center and ISAS — that was to have measured the heat created by X-rays rather than converting them into electrical charges, and a 'hard' X-ray detector, built jointly by scientists at ISAS and the University of Tokyo.

For Kazuo Makishima, a professor at the High Energy Astrophysics Laboratory at the University of Tokyo, who was involved in building the hard X-ray detector, the loss of Astro-E is a serious blow to X-ray astronomy in Japan.

Yasunori Matogawa of ISAS, the key person behind Japan's X-ray satellite programme, says the failure has come at a "a very bad moment" for Japan. "With XMM-Newton in Europe and Chandra in the [United States], we were

looking towards a period of exciting research and competition."

Japan should strive to launch a satellite similar to Astro-E as soon as possible, rather than trying to produce improved detectors, argues Matogawa. But other launch commitments make it unlikely that a new mission could take place before 2004.

Following last week's failure, leading Japanese politicians have repeated calls for the country's two space agencies, ISAS and the National Space Development Agency (NASDA), to be merged into a single organization.

Such a move was suggested a few months ago, when one of NASDA's H-II rockets was destroyed because of an engine failure.

Robert Triendl

Russian presidential favourite pledges support for science

Moscow

Vladimir Putin, Russia's acting president and prime minister, said last week that scientists "must get salaries higher than anyone else in the country".

Speaking during a visit to Zelenograd, a town outside Moscow that is known as Russia's 'Silicon Valley', Putin said that "if a country lacks a working economy, it has no present, but if it lacks a competitive science, it will never have a future".

The visit took place on 8 February, the date on which Peter the Great signed his decree founding the Russian Academy of Sciences in 1724, and which was designated last year as the annual Day of Russian Science by then president Boris Yeltsin.

In a bid to demonstrate that the new administration has a keen interest in science, many of the nation's top officials were present.

In an address, Putin said that the process of acknowledging the economic and strategic importance of science had already started. He pointed out that last year, for the first time in a decade, all the money allocated to science in the state budget had been paid.

He added that he was confident that this current year would see a breakthrough in funding for science.

Science provided the United States with 20 per cent of all its wealth, but Russia with less than 1 per cent. "The major problem here is an abyss between science and its implementation," said Putin. "As time passes, we are still unable to derive benefits from our own ideas."

Although stopping short of promising a substantial increase in the science budget, he suggested that the whole approach to raising money for research should be changed. It needed to be multichannel and to make effective use of intellectual property. Most important, said Putin, was that scientists should remember that the "real value of their research is confirmed by the market".

On the eve of his visit to Zelenograd, Putin met privately with Yuri Osipov, president of the Russian Academy of Sciences, a move which suggests that real reforms may be implemented. The planned science budget for the year 2000 is already 40 per cent higher than last year, with a high proportion of this being allocated to basic research. But the total science budget is just 15.9 billion rubles (\$550 million).

Carl Levitin

Republican candidates clash on fetal tissue in research

Washington

As George W. Bush, the governor of Texas, and Arizona senator John McCain prepare to contest the Republican primary in South Carolina on 19 February, the controversial issue of government-funded fetal-tissue research is emerging as a flashpoint in the race for the presidential nomination.

The National Right to Life Committee (NRLC), the nation's largest anti-abortion group, is running radio advertisements attacking McCain for reversing his opposition to government funding of such research. In a series of Senate votes between 1992 and 1997, McCain supported the research, even though he had opposed it up to 1992.

Bush is capitalizing on the issue. "Governor Bush is opposed to federal funding for medical research involving aborted fetuses," says Scott McClellan, Bush's press secretary. "It's up to Senator McCain to explain why he changed his position."

But supporters of the research charge that Bush and the NRLC are distorting McCain's position. "A vote on fetal-tissue research is not a pro-abortion vote," says Joan Samuelson, president of the Parkinson's Action Network.

McCain has said that he was moved to change his position by the plight of his close friend, the late Arizona Congressman Morris Udall, who died from Parkinson's disease.

"I've been convinced that [fetal-tissue research is] promising... a way to find a cure



At odds: McCain (right) takes a more liberal stance than his rival Bush.

for a terrible disease. And I made no secret of my changing my position," McCain said on ABC television's *This Week* on 6 February. "I'm not supporting abortion to provide it."

But in a letter written to an anti-abortion constituent in 1998, McCain said that "I am appalled, as you are, at the prospect of induced abortions becoming an accepted practice, even if it is in the interest of 'furthering medicine'".

Apart from the issue of fetal-tissue research, McCain's voting record has otherwise been strongly anti-abortion. For example, he was among 20 senators who signed a 4 February letter asking the National Institutes of Health to withdraw its proposal to fund research on human stem cells derived from embryos left over after fertility treatments.

McCain is among several high-profile Republican senators who oppose abortion but have supported federal funding of fetal-tissue research.

Meredith Wadman

Funding crisis for Indian biotech centre

New Delhi

An international centre set up in India 12 years ago by the United Nations Industrial Development Organization (UNIDO) to bring the benefits of biotechnology to developing countries is facing a financial crisis.

"The future is grim," says Virender Singh Chauhan, chief executive of the New Delhi component of the International Centre for Genetic Engineering and Biotechnology (ICGEB). Grants for his unit have been cut by 10 per cent this year, and support is expected to be reduced by about 25 per cent by 2001.

The ICGEB, whose rules and regulations have been ratified by the governments of 61 countries, operates from two locations, in New Delhi and at Trieste in Italy.

It had been hoped that the centre would raise money from members' contributions, grants from research foundations, the commercialization of technologies it develops, and support from India and Italy.

According to Chauhan, the recent budget cuts were forced by the member states refusing to pay their contributions.

Non-payment by member states has not affected the Trieste unit, as the Italian government has doubled its contribution, says Chauhan. The New Delhi unit is surviving on an annual contribution of US\$1.25 million from the Indian government and grants from ICGEB headquarters in Italy. But the grants have been cut, and the Indian government is not willing to increase its contribution.

The crisis has already taken its toll of research programmes. Recruitment has been halted, and 16 scientists have already left, with 10 more likely to follow.

Chauhan says that so far only middle-level researchers have left, but that senior scientists may leave if the situation does not improve. The number of scientists trained by ICGEB has also fallen as funding for training programmes been reduced, he says.

K. S. Jayaraman

Oxford professor faces business link inquiry

London

Both Oxford University and the Wellcome Trust are investigating the business affairs of Roy Anderson, the suspended zoology professor and Wellcome Trust governor.

The inquiries are in response to concerns raised during a separate inquiry into Anderson's behaviour in connection with the appointment of a faculty member in the department (see *Nature* **403**, 472, 353; 2000).

The investigations are believed to include Anderson's involvement with the company International Biomedical and Health Sciences Consortium (IBHSC), an Oxford-based biomedical consultancy. Anderson is one of the directors, and holds one-third of its shares.

Anderson is a governor of the Trust, a director of the Wellcome Trust Centre for the Epidemiology of Infectious Diseases, and a respected scientist and a government adviser known for his pioneering work on AIDS and on the spread of BSE among British cattle.

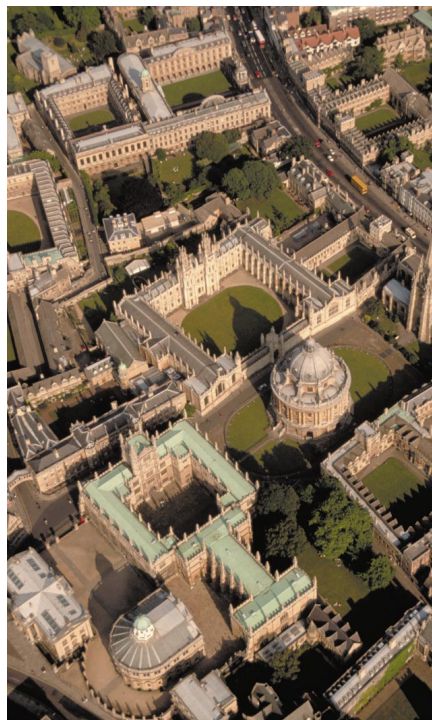
He is currently suspended from Oxford University and his position as director of the Centre, while the inquiry into the faculty appointment is underway. Both Oxford and the Wellcome Trust emphasize that their inquiries are neutral, and make no presumption of innocence or guilt.

The Trust and the university will cooperate on the new inquiry on matters of mutual interest. The Trust says that its inquiry will be completed in weeks rather than months, and has promised that there will be "no attempt to hide anything".

One of the Trust's guiding principles is that the research it funds is charitable and not constrained by any organization or grant holder's commercial interests. For example, Trust-funded researchers may not consult for a commercial organization in which they hold a significant equity interest, unless the consultancy relates to an area outside the Trust's interests.

The Trust generally defines a significant equity interest as more than 5 per cent of the organization's outstanding equity interest. Its rules add that, in some circumstances and at the discretion of the Trust, researchers may hold more than this after obtaining prior written agreement.

Anderson and Jack Crew, a business consultant, are directors of IBHSC, which is based



Oxford University is carrying out "inquiries within the Wellcome Trust Centre".

at the Oxford Science Park. IBHSC has client pharmaceutical companies in Europe and the United States, with contracts to work on the use, delivery, epidemiology and economics of anti-infectives and vaccines. A number of scientists from the Wellcome Trust Centre carry out contract work for IBHSC.

Mike Dexter, the director of the Wellcome Trust, has declined to confirm or deny that the Trust has agreed to Anderson's IBHSC activities. Asked whether Anderson's involvement with IBHSC was in line with Wellcome Trust rules, he also declined to comment. Dexter added: "Obviously we take any issues raised seriously but we must carry out due process."

In response to a question from *Nature* about the continued tenability of Anderson's position as governor, given the new investigation, the Trust said in a statement this week that it is "carrying out inquiries into the concerns raised by yourself and others. As these are currently underway it would be inappropriate for us to comment. We do not want to prejudice any issues."

Oxford University also confirmed last week that it is examining the links between Anderson, his company IBHSC and the drug company Abbott Laboratories. A spokesperson said "I can't say too much about this business other than to reassure you it has been drawn to our attention and will be investigated."

Later, an official statement said that the

university was carrying out "inquiries within the Wellcome Trust Centre in accordance with our standard practices. It would be inappropriate to prejudice the outcome of these inquiries in any way."

IBHSC was incorporated on 3 October 1997; accounts from this date up to 31 March 1999 show that IBHSC held £1.1 million cash at the bank and in hand. Anderson and Crew each hold exactly a third of the equity of IBHSC; the other shareholder is Camrep Investments SA, in County Cork, Ireland. Efforts to locate the address of the company or contact either it or its registered offices have been unsuccessful.

As well as being a governor of the Trust, Anderson is the principal applicant on three current Wellcome Trust awards valued at more than £4 million (US\$6.3 million).

The inquiries will be embarrassing for both the Trust and the Centre, which has acknowledged academic links with Abbott Laboratories. A recent paper on antigen-driven CD4⁺ T cell and HIV-1 dynamics (*Proc. Natl Acad. Sci. USA* **96**, 15167–15172; 1999), co-authored by researchers from the Centre, acknowledges Abbott for "advice and access to pharmacological data".

Abbott is a world leader in anti-infective research. Their innovations include the antibiotic Biaxin and Norvir (Ritonavir), an HIV protease inhibitor.

Anderson's Wellcome Trust grants are for the Centre's core funding, research into the population and evolutionary biology of infectious diseases, and a supercomputer for investigating the persistence, dynamics and control of human and animal infectious diseases.

In recent years, the Trust has expressed concern over the huge amount of work and responsibility Anderson has taken on, including work as a governor and director of the Centre.

This is said to have led to an attempt to persuade him to step down as a governor — which failed after Anderson's lawyers objected on his behalf — and to Anderson's decision last year not to take up the position of biological secretary of the Royal Society to which he had been elected.

Last week Anderson's lawyers said that they would respond "vigorously and quickly" to any suggestion of wrongdoing on Anderson's part. They declined a request for an interview with Anderson. In the past week Crew was also unavailable for interview, despite previously agreeing to be interviewed.

Abbott Laboratories were asked about their involvement with IBHSC and the Centre, but were unable to respond before *Nature* went to press.

Natasha Loder

Busquin set to create league tables for European research

Brussels The European Commission is to set up a benchmarking system to compare the scale and quality of research in the 15 member countries of the European Union. This annual league table of research effort — an initiative by research commissioner Philippe Busquin — will identify ‘centres of research excellence’, raising the prospect that future EU funding will be more sharply focused on results.

Busquin aims to create a ‘European research area’ to bring science and technology closer to the political decision-making process and to introduce a European dimension to science. This, he says, “will attract researchers from third countries and encourage the return of European researchers”.

So as well as recording how much a country spends on research, the league tables will show how open domestic programmes are to nationals of other EU countries, and the number of people — including non-nationals — who sit in research councils.

All public research will be covered, together with private research “to the best of our ability” says a spokesman for Busquin, conceding that the latter will depend on

companies’ goodwill. Busquin has stressed the importance of establishing quantitative performance criteria, such as the number of patents awarded or the number of researchers employed by companies.

Online bioscience repository is launched

Paris PubMed Central, which gives free access to the full text of life-science papers, went live this week (<http://www.pubmedcentral.nih.gov/index.html>). The initial offerings are papers from the December issue of the *Proceedings of the National Academy of Sciences* and the November issue of *Molecular Biology of the Cell*. The *Biochemical Journal*, the *Canadian Medical Association Journal*, *Frontiers in Bioscience* and BioMed Central will be available soon.

Attention on free repositories has focused on PubMed Central and its proposed European counterpart, E-Biosci. But the website of HighWire Press (<http://highwire.stanford.edu/lists/largest.dtl>) — a non-profit organization set up in 1995 by Stanford University Libraries and Academic Information Resources — reminds readers that, with 134,143 full-text articles, it is the second-largest repository, after the US space agency NASA’s Astrophysics Data System, which has more than 300,000 free articles.

Biotech company’s share price soars

Washington Curagen, a biotechnology company based in New Haven, Connecticut, which specializes in genomics-based drug discovery, is the latest beneficiary of the boom in genomics stocks. Last week its share price almost doubled — rising from \$91 to \$175 — with the publication of a paper in *Nature* describing its use of sequence data to analyse interactions between yeast proteins (see *Nature* **403**, 623–627; 2000).

Some analysts noted that there was a rush to buy shares in the company two days before publication, suggesting that this may have been prompted by a leak of information on an Internet noticeboard. Company officials decline to discuss the alleged leak, saying that they are forbidden to comment on the company’s share price.

US collaboration to develop malaria vaccine

Washington The US National Institute of Allergy and Infectious Diseases (NIAID) agreed last week to collaborate with the non-profit Malaria Vaccine Initiative (MVI) of Rockville, Maryland, on developing a malaria vaccine.

The Seattle-based Bill and Melinda Gates Foundation established the MVI last year

with a \$50-million grant. The organization will channel funds into NIAID's budget for research activities agreed by the organization and the government agency.

US ban on misuse of genetic information

Washington US president Bill Clinton last week signed an executive order forbidding any US federal agency from using genetic information as a basis for hiring, promotion or dismissal of staff. The president also endorsed the Genetic Nondiscrimination in Health Insurance and Employment Act of 1999, introduced by Senator Thomas Daschle (Democrat, South Dakota) and Congresswoman Louise Slaughter (Democrat, New York), which would extend these protections to the private sector and to people purchasing health insurance.

Japan sets up expert panel on impact of GM crops

Tokyo The Japanese Ministry of Agriculture, Forestry, and Fisheries (MAFF) has announced its intention to set up a special expert panel under the Agriculture, Forestry, and Fisheries Research Council to deal with the potential environmental risks associated with genetically modified (GM) crops.

The panel will be asked to discuss adequate safety measures for GM crops and to define policies for research on the environmental impact of GM organisms which are sensitive to mounting societal concern. In addition to experts in plant genetics, ecology and risk assessment, the committee will include representatives from the social sciences and from the media.

Commission backs dismissal claim

Montreal An investigation by the Ontario Human Rights Commission has found prima facie evidence that "race, colour, ancestry, place of origin and ethnic origin" were factors in the failure of a physicist, Kin-Yip Chun, to obtain an academic appointment at the University of Toronto, and that "he was subjected to a series of reprisals culminating in his dismissal". It has given the university three weeks to respond and recommends an inquiry.

Chun was hired by the university in 1985 as a research associate for the physics department and subsequently entered four separate competitions for tenure-stream positions of assistant professor (see *Nature* 392, 638; 1998). He alleges that he was rejected in favour of less-qualified white applicants, and was dismissed in 1994. The university says that the four competitions

were conducted in accord with university policy, and that his appointment was not renewed partly because of his "irrational behaviour".

Tributes follow death of Australian science adviser

Sidney Malcolm McIntosh, chief executive of Australia's large research agency, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), died last week at the age of 54, two days before he was due to be a leading player in the nation's first Innovation Summit.

Acknowledged as an influential adviser on science policy to both Labor and Liberal governments, McIntosh's energy was legendary, especially given the regular dialysis he needed after losing both kidneys to cancer before returning to

Australia from Britain in 1996.

Science minister Nick Minchin last week praised "his indomitable spirit, his intellect and his great zest for life" and named a new Young Australian Science Award after him.



CSIRO

Reliance on the citation index undermines the study of biodiversity

Sir — The popularity of the Science Citation Index (SCI) as a measure of 'good' science is damaging basic taxonomic work, without which the study of biodiversity would not be possible.

Basic taxonomic work is not highly cited, except in 'hot' taxa like the genus *Homo*. The number of authors citing a paper during the short period of time (ten years) that the SCI uses for its statistics is relatively low. But taxonomy papers continue to be referred to and cited for more than a century after their publication. Almost every good taxonomic paper becomes a classic in the literature.

High-quality basic taxonomic work — the description of new taxa and revision of older ones — is expensive and time-consuming. Many of the most interesting finds are from 'exotic' locations, requiring travelling, sampling, preparing, sending back collections, writing descriptions, illustrating and so on. The resulting papers are rated low in the SCI, even when published in high-quality specialist journals, and are unlikely to impress managers or funding agencies.

So a paradox arises: concern for biodiversity goes together with a dismissal of the foundation of any biodiversity work, which is the proper description of taxa. If there is reluctance to fund this kind of work because of low citations, and with fewer journals available to publish their findings, the most basic research in biodiversity is doomed to disappear, as is already happening.

Antonio G. Valdecasas*,

Santiago Castroviejo†, Leslie F. Marcus‡

**Museo Nacional de Ciencias Naturales (CSIC), José Gutiérrez Abascal 2, 28006 Madrid, Spain*

†*Real Jardín Botánico (CSIC), Plaza de Murillo 2, 28014 Madrid, Spain*

‡*Department of Biology, Queens College of City University of New York, Flushing, New York 11367, USA*

Proteomics could be key in battle against malaria

Sir — A human proteome project may be premature (*Nature* **402**, 703; 1999), but cataloguing proteins of life-threatening single-celled organisms such as the human malaria parasite *Plasmodium falciparum*, and another intracellular parasite, *Leishmania*, will be interesting, as genome sequencing of these parasites is close to completion.

Understanding the activity of all proteins will be "frustratingly tantalizing" for now, but investigators cannot wait for the advanced technologies they envisage. The creative imagination and ingenuity of the human mind will take up the challenge by employing tomorrow's technologies today, on a limited scale. With these parasites, even two-dimensional protein profiles will be useful in exploring some important issues — investigating the elusive chloroquine-resistance mechanism, for instance, by comparing two-dimensional gels of chloroquine-sensitive and -resistant clones of *P. falciparum*.

Employing tricks such as zoom gels, it may be possible to identify the protein or proteins involved in imparting drug resistance to these parasites. Identification of the drug-resistant/sensitive protein markers will provide a lead in developing new antimalarial drugs or reviving chloroquine efficacy.

The proteomics conducted on these organisms should be carried out on a large scale. It has much to offer and requires the immediate attention of the developed world — after all, it is a matter of tropical health in a global village.

Virendra K. Bhasin

Department of Zoology, University of Delhi, Delhi 110007, India

X chromosomes forget where they came from

Sir — In an excellent News and Views Feature on the future of somatic nuclear-transfer cloning, J. B. Gurdon and Alan Colman (*Nature* **402**, 743; 1999) express concern about the inactive X chromosome when female cells are used in cloning. If the inactive X chromosome does not become reactivated by exposure to egg cytoplasm, they wonder whether embryos reconstructed from a donor containing an inactive maternal X chromosome be viable.

On this point they need have few worries. In early embryos of eutherian mammals, the X chromosome derived from the father is indeed distinguished from that coming from the mother, and is preferentially inactivated in the first two tissues to differentiate (trophoblast and primary endoderm, both of which contribute exclusively to the placenta and extra-embryonic membranes). Thereafter, however, the X chromosomes 'forget' which parent they have come from: they are inactivated at random in the cells contributing to the fetus and retain no trace of their parental origin.

Incidentally, it is legal in the United Kingdom to use human eggs to create an

embryo, if the aim is to satisfy one of the purposes specified in the 1990 Human Fertilization and Embryology Act (which does not include research on cell or tissue therapy). An amendment seeking to prohibit the use of human eggs to create embryos for research was put forward during the passage of the Act, but was soundly defeated in both Houses of Parliament.

Anne McLaren

Wellcome/Cancer Research Campaign Institute, Tennis Court Road, Cambridge CB2 1QR, UK

Don't let politics put Diamond at risk

Sir — It is with increasing dismay that I watch the antics over the siting of the new UK synchrotron, Diamond (see *Nature* **402**, 451; 1999). The UK government now appears to be in dispute with the Wellcome Trust — the world's biggest medical charity. This has resulted in further delay to a facility which, in scientific terms, is overdue, as well as in harsh criticism of the Wellcome Trust by politicians and scientists.

Lest we forget, it is the Wellcome Trust, not the government, that has saved UK biomedical science: without its financial support, cuts to science funding would have been fatal. Do we seriously believe that the government (or voters) will put structural biology above tax cuts or hospital beds?

I think an open competition would have been the best way to decide on the siting of Diamond. But the Kafkaesque behaviour of the Office of Science and Technology has ensured that this is no longer an option. The choice is now between Daresbury in the north of England or Rutherford in the south.

My own preference would be for Daresbury. But if the cost of siting Diamond there is a loss of Wellcome involvement, then the price is too high. Wellcome is vital to the project, not only in cash terms (building a smaller synchrotron ring would be worse than useless), but also in contributing dynamic scientific and intellectual management expertise.

The United Kingdom has a chance to build a world-class facility in structural genomics to complement its world-class facilities in genome sequencing and bioinformatics. Both these rely on substantial support from the Wellcome Trust. We must have Diamond and we must have it soon — to lose or degrade such a precious jewel would be a disaster.

James H. Naismith

Centre for Biomolecular Sciences, University of St Andrews, Fife KY16 9ST, UK

One-stop shop for microarray data

Is a universal, public DNA-microarray database a realistic goal?

**Alvis Brazma, Alan Robinson,
Graham Cameron
and Michael Ashburner**

Of the techniques that are being used to obtain the massive data sets of the molecules of life, the most visible is the DNA sequencing of the human genome. Following on from the publication of the human chromosome 22 sequence¹, a rough draft of the whole human genome should be available by the spring. But such advances can create the false impression that everything about life at the molecular level will soon be understood.

In reality, genome projects simply transfer digital information from DNA to computer file; this genetic 'parts-list' is a long way from providing an understanding of function. It took hundreds of years to advance from a fairly detailed understanding of human anatomy to any real understanding of function. Knowing the genome sequence and even the location of all an organism's genes is the 'anatomical' description of its genome. Functional genomics is the science of understanding how the genome functions through controlling the expression of genes. This scientific discipline will continue long after the DNA sequence of many genomes is known. To gather information about gene expression is a much more formidable task than DNA sequencing, because gene expression is a dynamic process dependent on many factors.

With the advent of DNA microarray and 'chip' technologies (see Box 1), gene expression in an organism can be examined on a genomic scale, allowing the transcription levels of many genes to be measured simultaneously². For instance, we can study the effects of a compound (such as a drug) on the level of expression of many genes. The pattern of genes expressed, say, in human brain and muscle, are very different, and will vary with the time and conditions under which samples are taken — before or after exercise, or perhaps before or after listening to Verdi. Yet each cell in the brain and muscle is using the same genomic sequence to regulate its function. With gene expression, context is everything: without it, the information is meaningless. For example, the precise stage of a tumour sample could have a crucial bearing on the interpretation of expression measurements. This context can be infinitely detailed, and it is this detail that must be captured in gene-expression studies.

The bioinformatics underlying the management of these huge volumes of data are crucial if any sense is to be made of gene-expression experiments. A single microarray

experiment looking at 40,000 genes from 10 different samples, under 20 different conditions, produces at least 8,000,000 pieces of information. Currently, these data are scattered among various independent Internet sites, or may not be publicly available at all, although conclusions drawn from the data will have been published. Details about how experiments were carried out are often incomplete. Yet the amount of information being produced in this way is set to explode as the cost of microarray technology falls.

The need for a public repository

It is time to create a public repository for microarray data, with standardized annotation (see Box 2, overleaf). But this is a complex and ambitious project, and is one of the biggest challenges that bioinformatics has yet faced. Major difficulties stem from the detail required to describe the conditions of an experiment, and the relative and imprecise nature of measurements of expression levels. The potentially huge volume of data only adds to these difficulties. However, it is this very complexity that makes an organized repository necessary.

Important tasks to be undertaken include: (1) agreement on the essential information that should be reported for a microarray experiment; (2) definition of ontologies and an extensible, structured document format to capture these data and their semantics; (3) production of a database to store these documents; and (4) development of tools for searching documents in a database and using the semantic context to allow comparisons and sophisticated queries.

One difficulty concerns the inherent fuzziness of gene-expression data. Essentially all current expression measurements are relative: we can tell which genes are expressed differently in an experiment only in comparison with another experiment, or in relation to another gene in the same experiment. Such methods tell us little about how many copies of a messenger RNA are present. Moreover, the transcription levels reported are an average over the whole cell population sampled.

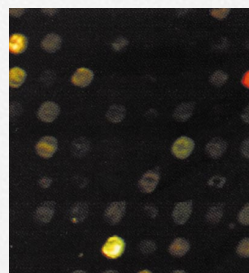
Consequently, gene-expression measurements from different technologies, or even from the same technology but from different laboratories, may not be quantitatively comparable. Two steps should allow data from different sources to be compared. First, relatively raw data should be stored to obviate any variation owing to, say, data-normalization methods. Second, standard sets of control probes and samples should be designed and used in experiments to give reference points so that these data can be normalized, at least from the same experimental platform.

The ability to compare results obtained using different technologies will depend on careful comparison of the technologies; such experiments should be encouraged and published. Realizing the full potential of chip methodologies will depend on basic research of this type, which documents the behaviour of the technology rather than advancing biological knowledge.

Initial solutions to the second difficulty — describing experimental conditions — will be a mixture of the formal and the pragmatic. Some parameters will be obvious and may be described quite precisely, probably by using

Box 1: How microarrays work

DNA microarrays exploit the preferential binding of complementary, single-stranded nucleic-acid sequences. A microarray is a glass slide, to which single-stranded DNA molecules are attached at fixed locations (spots). There may be tens of thousands of spots on an array, each representing a single gene. RNA from the sample and from control cells is extracted and labelled with two fluorescent labels: for example, a red dye for RNA from the sample population and a green dye for that from the



control population. Both extracts are washed over the microarray.

Gene sequences from the extracts hybridize to their complementary sequences in the spots. The dyes enable the amount of sample bound to a

spot to be measured from the level of fluorescence emitted when it is excited by a laser. If the RNA from the sample population is in abundance, the spot will be red; if the RNA from the control population is in abundance, it will be green; if sample and control bind equally, the spot will be yellow; if neither binds, it will appear black. Thus, the relative expression levels of the genes in the sample and in control populations can be estimated from the fluorescence intensities and colours for each spot.

documented experimental protocols and controlled vocabularies: for example, for gene names, organisms, developmental stages, tissue or cell lines (although most of these vocabularies have still to be developed; see <http://www.geneontology.org/>). Creative experimentation will explore how gene expression responds to a range of treatments and exposure to various compounds, and expression may be measured over a time series. However, incidental and uncontrolled aspects of experimental conditions may also affect expression: for example, the time of day, air humidity, or even the noise level in the laboratory. To build a formal structure describing this would be to anticipate all the niceties of future experimental design — clearly impossible. A realistic solution is to describe experiments largely in free text, while imposing as much structure as practicable.

To solve these problems, standards for annotating microarray experiments must be based on agreed ontologies. These will take time to develop; right now, biologists should agree on as many sensible standards for microarray data and annotation as possible, and we should build a flexible data model that can evolve with our understanding of gene-expression experiments.

Structured information storage

Fortuitously, many problems in the storage and description of gene-expression data and experiments parallel current problems with the Internet — too much information, which is hard to search without being deluged by irrelevant information. One reason is that HTML, the language used for documents on the Internet, cannot describe any semantics of the information it carries — hence the development of XML (Extensible Markup Language)³ for defining data formats for storing semi-structured and structured information. XML has been designed specifically to allow people to create their own customized mark-up languages for storing any type of data and their semantic content.

XML seems well suited as a format for the formal description and annotation of microarray information, and ways to describe microarray experiments using this language are being developed⁴. This will mark an important milestone in the formalization of standards for microarray data and facilitate exchange of information among researchers. XML specifications have already been proposed for genome annotation, biomolecular sequences and computational analyses⁵.

Several of the largest microarray laboratories in academic institutions — Stanford University and the Whitehead Institute, among others — are developing their own internal databases, as are the larger commercial microarray companies such as Affymetrix and Incyte. GeneLogic is creating a commercial gene-expression database using data obtained on the Affymetrix platform. Harvard

Box2: Uses of a microarray database

● **In basic research.** The integration of expression data from various organisms with data from other molecular-biology data resources, such as databases of annotated genome sequences, gene promoters and metabolic pathways, will enable us to build up an understanding of gene regulatory networks, metabolic pathways, cell differentiation and tissue development. For instance, similarities in gene-expression profiles will allow hypotheses on the roles of unknown genes to be formulated by comparing them with known genes.

● **In medical research and pharmacology.** For example, if overexpression of certain genes is correlated with a certain cancer, other conditions

affecting the expression of these or other genes that have similar expression profiles, or compounds (potential drugs) that lower the levels of expression, can be explored.

● **In comparative and diagnostic studies.** How does a gene-expression profile for a sample tissue compare with profiles for other samples in the database? The ability to detect a disease at an early stage through similarities in expression profiles would have major implications for health care.

● **In toxicology.** For instance, knowing the response of a gene to a certain toxin in rats will allow users to predict the response of a homologous human gene. Many

multinational pharmaceutical companies, coordinated by the International Life Science Institute, are collaborating to produce extensive gene-expression profiling for toxic responses. The need for a well-curated database for this project was recognized from the start.

● **In experimental methodologies and peer-review work.** Results of archived control samples can be compared with those from laboratory samples to check protocols and equipment for consistency and quality. The work of other researchers can be cited, validated and extended. Storing results from common systems avoids unnecessary repetition of experiments, and saves money.

University has created a database that compiles and represents data from several public sources in a consistent format. The Computational Biology and Informatics Laboratory of the University of Pennsylvania is incorporating ontologies to describe samples.

There are at least three projects for large, public gene-expression data repositories^{6,7}: GeneX at the US National Center for Genome Resources; the Gene Expression Omnibus at the US National Center for Biotechnology Information; and ArrayExpress at the European Bioinformatics Institute (EBI). Compatibility between these projects requires common standards for data representation, annotation and exchange, and discussions between these bodies are thus being focused on XML.

Working groups

Progress towards such standards was made last November, when many of the main academic and commercial users and developers of microarray technology accepted a list of recommendations for data representation and annotation (see <http://www.ebi.ac.uk/microarray/>). Five working groups, coordinated by US and European specialists, have been set up to develop standards in experiment description and data representation; microarray data XML exchange format; ontologies for sample description; normalization, quality control and cross-platform comparison; and data-query language and data-mining approaches. The users will meet again in May to discuss detailed recommendations from the working groups.

Meanwhile, the EBI, in collaboration with the German Cancer Research Centre, is

developing ArrayExpress, a gene-expression database compliant with the current recommendations. The first draft of the data model used in ArrayExpress was posted on the Internet in October (see <http://www.ebi.ac.uk/arrayexpress/>). This model will be refined in the light of discussions by working groups, and the next draft will be released in a few months' time, together with a database schema. The schema will be implemented with data from collaborating institutions, creating a prototype database by the spring. But it is a giant leap from a prototype to a working database, with associated tools for submitting and retrieving microarray data.

The estimated costs of setting up the permanent database will be more than 500,000 euros (US\$485,000) in the first year. The EBI's parent organization, the European Molecular Biology Laboratory, expects to be substantially increasing its investment in bioinformatics, and sees the development of ArrayExpress as a high priority. A fully operational database will be released by the end of the year. Steady-state costs will be greatly above these start-up figures, though somewhat dependent on the volume of data. ■

Alvis Brazma, Alan Robinson, Graham Cameron and Michael Ashburner are at the European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. e-mail: brazma@ebi.ac.uk

1. Dunham, I. *et al.* *Nature* **402**, 489–495 (1999).
2. The Chipping Forecast *Nature Genet.* **21**, suppl. (1999).
3. The World Wide Web Consortium <http://www.w3c.org/XML/>
4. *Nature Genet.* **22**, 211–212 (1999).
5. The BioXML Project <http://www.bioxml.org/>
6. Pennisi, E. *Science* **286**, 449 (1999).
7. Kaiser, J. (ed.) *Science* **286**, 1047 (1999).

Balancing ambition and restraint

India's struggle with its twin desires to have and to renounce the bomb.

India's Nuclear Bomb: The Impact on Global Proliferation

By George Perkovich

University of California Press: 1999. 597 pp.

\$39.95, £24.50

Robert W. Cahn

Nuclear technology is complex: extensive expertise is needed to design a safe nuclear reactor, to produce pure isotopes or to create an 'efficient' nuclear explosion. How much (and what kind of) diplomatic expertise is required to dissuade a nation from acquiring and exploiting nuclear technological know-how is another issue altogether. George Perkovich, a member of the US think-tank Secure World Program, is a social scientist and specialist in foreign affairs. In the words on the dust-jacket, he has "grappled with [India's] twin desires to have and to renounce the bomb". His central theme is the unceasing minuet between successive Indian governments and US politicians and officials engaged in anti-proliferation activism, a minuet that has been in progress for the past half-century.

This long, detailed chronicle, scrupulously pinned to sources throughout, covers the Indian experience from 1948, when Jawaharlal Nehru put Homi Bhabha in charge of the nascent nuclear effort in India, until the aftermath of the 1998 nuclear bomb tests. Perkovich judges that "India is an extremely important case to examine. The nation's size, potential in the international system, democratic system of governance, place in the Sino-Pakistani-India security triangle, moral traditions and other features give it exceptional and practical import."

But Perkovich is chary of generalization, although he constantly contrives to illuminate his themes. Since the Indian political system provides very few relevant governmental records (and those that exist record decisions but not how they were reached), he relies on remarkably varied interviews with past and present Indian and US officials — only a few of them scientists, engineers or military men — and on comments in Indian newspapers and books. In his acknowledgments, he refers to having "benefited greatly from weeks of stimulating, sometimes mutually exasperating conversations in India". The picture Perkovich builds up carries conviction, in spite of the difficulties of his *modus operandi*.

The single most important fact to emerge here, a fact insufficiently recognized, is that the Indian scientific and technological élite involved with nuclear weaponry and missile technology (an élite he calls the strategic



Cautious steps: Indian Prime Minister Indira Gandhi visits the nuclear test site at Pokhran in 1974.

enclave) was, for many years before 1974 and again before 1998 — the years of the two sets of nuclear tests at Pokhran — fully capable of creating such explosions, but was repeatedly restrained by cautious prime ministers and, for a time, by an Atomic Energy Commission chairman, the eminent physicist Vikram Sarabhai, who is depicted here as "a bomb agnostic if not a sceptic". Despite considerable alarm about China's and Pakistan's capabilities, genuine moral considerations held Indian leaders back from tests and, even more pressingly, from sanctioning large amounts of money on delivery systems (which US experience shows to be even more expensive than the weapons themselves). In spite of Bhabha's absurdly low estimates, in the early days, of the prospective costs of nuclear explosive devices (whether for peaceful or hostile purposes), Nehru and his daughter, Indira Gandhi, for a long time held back from sanctioning explicit sums.

The leaders of the strategic enclave, much more than the military, who, until recently, were kept out of the decision-making circuit, exerted steady pressure for weapons-related research, and later for work on missiles. Enough money was quietly made available

for background research; the scientists, for their part, avoided inflammatory utterances and succeeded in maintaining unity and high morale. There is a clear distinction here from the Pakistani situation, where the weapons establishment was split into two, with the leader of one faction constantly making boastful and threatening speeches. Further, to quote Perkovich, "the military role meant that Pakistan's nuclear weapon capability would be ingrained in military planning much more deeply than India's had been or would be". Although China carried out many nuclear tests over a long period, her demonstrated nuclear capabilities always caused less alarm in Delhi than did Pakistan's presumed activities.

Four political/economic factors loom large in the book. First, the intimate link between nuclear armaments and general economic success: "The need to drastically improve Indian industry and infrastructure pointed to the profound flaw in Bhabha's initial plan: it failed to recognize the nuclear programme's organic dependence on the broader national economy." At intervals over the half-century covered by the book, finance ministers warned that even the

understated cost estimates from the strategic enclave might be disastrous for the fragile Indian economy.

The second factor is the Indian passion for non-alignment between the two sides of the Cold War, sometimes more honoured in the breach than the observance, but always loudly declared in the Lok Sabha — India's parliament — and linked with an emphasis on the need for India to behave 'morally'. This emphasis on morality, going back to Mahatma Gandhi's teaching, was a key aspect of Nehru's character.

The third aspect that runs through the book is "the hugely important problem of esteem. ... How could India sustain its self-image as a great civilization and nation worthy of a leading place in the international system if America treated it as inconsequential?" The cynical behaviour of Henry Kissinger towards India when he and President Richard Nixon were seeking their opening to China in 1970–71 bedevilled India's response to US diplomacy for many years. India has always been ultra-sensitive to quasi-colonial attitudes, and its leaders detected these in US behaviour (more than that of Britain). The added fact that the United States for years felt constrained to ignore Pakistan's nuclear preparations because of the 'opening to China', and the US need for Pakistan's help against the Soviet forces in Afghanistan, compounded India's almost paranoid attitude towards the United States' condescending posture.

Linked with the issue of esteem is the book's fourth topic, which is the US determination, over many years, to impose the Nuclear Non-Proliferation Treaty on India. For a long time, this amounted to all stick and no carrot, and India's elite understandably saw no reason to sign away their future freedom of action as long as the five 'authorized' powers did nothing to reduce their own nuclear arsenals. India was well aware of the likelihood of painful economic sanctions should it cross the nuclear threshold (the central 'stick' of US diplomacy), but the total failure of successive US administrations to offer any sort of 'nuclear umbrella' as a quid pro quo for signing the treaty rankled deeply. Only recently did US carrots begin to be in evidence, and by then the Hindu vegetarian appetite had attenuated somewhat.

In its coverage of Indo-US relations, the book makes no mention of the numerous occasions on which Indians and Pakistanis talked at successive meetings of the Pugwash movement. Neither is there anything about British links with India, although British governments have clearly always been deeply concerned about nuclear proliferation. Nor is there anything about the strong influence on India of Patrick Blackett, the British Nobel-prizewinning physicist, military expert and president of the Royal Society. And there is little about science here, the only exception

being some interesting pages devoted to the distinction between weapons-grade and reactor-grade mixes of plutonium isotopes, and some remarks about neutron initiators. As a consequence, the book should be read in conjunction with *Atomic Energy in India: 50 Years* by C. V. Sundaram, L. V. Krishnan and T. S. Iyengar (published privately by the Indian Atomic Energy Commission in 1998).

Nevertheless, what Perkovich's book lacks fades into insignificance compared with what it achieves, the result of immense labour and great political expertise. It is essential reading for those concerned with the issue of non-proliferation. ■

Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

Between heavens and cloister

Galileo's Daughter

by Dava Sobel

Fourth Estate/Walker: 1999. 429 pp.
£16.99/\$27

Brenda Maddox

This is a harrowing tale of two victims: Galileo Galilei (1564–1642), forced to deny the evidence of his telescopes, and his daughter Virginia (1600–34), forced at the age of 13 to enter the living sepulchre of the cloistered convent. By entwining the two tales, Dava Sobel's latest well-timed bestseller provides a woman's angle on history's greatest conflict between science and religion.

This double perspective also makes sense of what can seem incomprehensible today: how Galileo, the father of modern physics, could have disclaimed his knowledge from observation that the Earth revolves around the Sun. Now it is clear that he recanted not only to save his skin but also because his devoted daughter — whom he called "a woman of exquisite mind" — wanted him alive.

Once more, as in her *Longitude* (Fourth Estate) of 1996, Sobel uses astute scholarship and scientific literacy to unearth a hidden drama. In the National Central Library of Florence, she found 124 letters written by Suor Maria Celeste to her father. The letters' existence was not unknown, but they had never before been thought worthy of translation into English. By doing

A victim of her time: but her father called her 'a woman of exquisite mind'.

this, Sobel reveals a young woman of unusual intelligence, scientific understanding and filial love.

From behind the walls of the Convent of San Matteo just outside Florence, where her younger sister was also sequestered, Suor Maria Celeste (born Virginia Galileo) was able to give her beleaguered and ailing father love and support during his trial and conviction by the Holy Office of the Inquisition in Rome in 1633. Devout Catholics both, father and daughter jointly believed that there was no conflict between religious faith and his astronomical discoveries.

Virginia took her religious name, Maria Celeste, from her father's absorption with the heavens. In her cloister, she read his *Book of Floating Bodies*, which contradicted the Aristotelian dictum that ice is heavier than water, and his *Dialogue on the Two Chief Systems of the World: Ptolemaic and Copernican*, which got him into trouble with the Inquisition. From the convent, too, through messengers, she looked after her father's parlous health, sending him herbal remedies and treats of food. She attended to his laundry, embroidering his linen and making his collars.

Of the two victims, the daughter suffered worse. Galileo got off lightly. Although the clerical censors found his *Dialogue*, even though he had craftily set it in fictional form, a scandalous glorification of Copernicus, he was not burned at the stake as Giordano Bruno had been in 1600 for defending heliocentrism.

Instead, the *Dialogue*, which he had had the audacity to write in the vernacular Italian, rather than Latin, survived and circulated. In 1635 the banned book was smuggled to Strasbourg, read throughout Europe and translated into English in 1661. In 1835 it was removed from the Church's Index of prohibited books. And in 1992 the current Pope removed whatever shadow remained on Galileo's Catholic reputation by forgiving both sides for mutual misunderstanding.

Sobel makes clear that it was the Office of the Inquisition rather than the Church itself that found Galileo guilty, and notes that papal infallibility, a dogma not formally defined until the nineteenth century, was never invoked in the conviction against Galileo.

Suor Maria Celeste was less fortunate. She died in 1634 at the age of 33, of dysentery and malnutrition, shortly after her father was released from detention in Siena and able once more to visit her. She lived for these visits, but all contact with visitors had to take place through a grilled window. Her letters asking her father's help in replacing



the convent's unsympathetic confessor-priest give some hint of the suppressed passions that raged under the black veil and rough brown linen habit of the Franciscan order of the Poor Clares. Indeed, the regime of the community — the sisters were perpetually barefoot, forced to sleep on bare boards and woken at midnight for hours of prayer — are harsh enough for any Ulsterman's nightmare of Rome.

One hard fact above all shines from this book. Until the twentieth century, the female revolved around the male. Women did what they were told, even if it meant, as in the case of both of Galileo's daughters, being shorn of their hair in pre-adolescence and sequestered for life to accommodate their father's social and economic circumstances.

Galileo's reasoning was clear. His children were illegitimate. Living for many years as professor of mathematics at the University of Padua, he never married his young mistress, the mother of his three children, because she was of lower social station. Burdened already by the responsibility of finding a dowry for his sister, he could offer none to his two daughters. For the unmarriedable, the convent offered the only economic recourse. Galileo, nonetheless, did persuade his Medici patron to legitimize his 13-year-old son, Vincenzo Galilei, a favour sweetened by the grant to the boy of a pension for life. Vincenzo went on to law school and marriage to a young woman with a substantial dowry.

The chance to explain Galileo's work brings out Sobel's science-writing best. She explains how Galileo perfected the Dutch spyglass, and with his telescope discovered the moons of Jupiter and the mountains of the Moon, and how, in 1613, he obediently put aside astronomy for eight years after the Holy Fathers passed an edict against anyone denying the scriptural truth that the heavens are immutable and the Earth is at the centre of the Universe. Her mechanical descriptions are as felicitous for Galileo's compass as for the clocks of the eighteenth-century Joseph Harrison in *Longitude*.

She is excellent, too, on the history of the Europe of the Thirty Years' War, and of the bubonic plague that swept the continent, isolating city from city, invading Italy from Germany in 1629. When Sobel comes to the unworldly, gifted daughter, however, a pofaced reverence takes over. A certain stiltedness was perhaps inevitable, as Maria Celeste's letters are couched in effusive phrases such as "Most Illustrious Lord Father", which could apply either to the scientist or the Saviour.

The book's design unfortunately contrives for a faux prayer-book tone. The preciousness of using Suor Maria Celeste's enlarged sloping Renaissance script for chapter headings is exacerbated by the author's indulgence in churchy archaisms, such as "in the Year of Our Lord" to specify

certain dates. There is also an embarrassing dedication, in which the name of Galileo Galilei is coupled with that of the author's own father, Samuel Hillel Sobel, MD.

These irritants, however, are secondary to a powerful reminder that the past is another country, in which people think and behave under conditions unimaginable today. The sea of faith, as Matthew Arnold reminded the increasingly rationalist late-nineteenth century, was once full at the flood. Religion held together the moral universe of Galileo and his daughter. From the evidence presented in this scrupulous contribution to the history of science, neither felt a trace of anger against the system that governed their minds and bodies. ■

Brenda Maddox is at 9 Pitt Street, London W8 4NX, UK.

A rival for the Burgess Shale

The Chengjiang Fauna: Exceptionally Well-Preserved Animals from 530 Million Years Ago

by Hou Xianguang, Jan Bergström, Wang Haifeng, Feng Xianghong and Chen Ailin
Yunnan Science and Technology Press: 1999. 170 pp. £40, \$70 (pbk)

Douglas Palmer

More than 120 species of fossils, including the oldest-known vertebrates recently described by D-G. Shu *et al.*, have already been described from Chengjiang in Yunnan Province, southwest China. The rock strata of this 530-million-year-old, early-Cambrian marine deposit are set to rival the Burgess Shale, British Columbia's World Heritage Site of similar age, which has yielded around 150 species.

Two-hundred or so colour photographs illustrate 12 new species and 72 other members of *The Chengjiang Fauna*. Altogether, they are one of the world's most important hoards of fossils. They open a fascinating window on the remarkable diversity of early Cambrian animals and their adaptation to an unexpectedly wide variety of niches close to a level seabed.

Although the text is in Chinese, there is a brief English summary and very useful bibliography to the primary literature. Hou Xianguang, the senior author, first found the deposit in 1984. As he records, it was 3 o'clock on the afternoon of 1 July when he broke open a piece of mudstone to reveal a beautifully preserved specimen of the arthropod *Naraoia*, previously known only from the Burgess Shale.

Interestingly, Cambrian strata and fossils



A window on diversity: *Naraoia longicaudata*, one of Chengjiang's rich hoard of fossils.

from this province of China were first found by the French geologists Jacques Deprat and Henri Mansuy in 1916. But they did not find the Chengjiang deposit. If they had, the débâcle over 'planted' fossils, which led to Deprat's dismissal from the Geological Survey of Indochina in 1920, may never have happened (see *Nature* 389, 688; 1997). ■

Douglas Palmer is at 31 Mawson Road, Cambridge CB1 2DZ, UK.

The book can be ordered from Hou Xianguang, Nanjing Institute of Geology & Palaeontology, Academia Sinica, Chi-Ming-Ssu, Nanjing 210008, People's Republic of China (£10 (\$17) postage & packing).

Problem squares and virtuous circles

A History of Algorithms: From the Pebble to the Microchip

edited by Jean-Luc Chabert
Springer: 1999. 524 pp. \$59.95, £39.95 (pbk)
Jeremy Gray

There is, at times, a deep division of intellectual labour in science. Some people formulate problems in the conceptual terms of a particular theory, some express the problems in mathematical terms, and some people solve them. It may be that one person does all three things, or that three protagonists are needed, none of whom communicates fully with the others. Each may nurture their own expertise while downplaying that of the others. In productive subjects, the solutions generate further

problems and the exchange forms a virtuous circle. This rich book, which reaches, as its subtitle indicates, from prehistory to the present day, illuminates the problem-solving side — the production of precise numbers, tables and functions.

Producing solutions can be challenging in a variety of ways. Sometimes a quick, accurate method is required, as in long multiplication. The defeat of the abacus users by those who favoured calculation with

numbers is an oft-told tale, and it is told again here. Another example is the inversion of matrices and the solution of linear equations. This topic was taken up by Karl Friedrich Gauss and, despite the tremendous speed of modern computers, it remains an issue because of the size of the problems we wish to tackle, such as weather forecasting.

Problems involving squares became, centuries later, quadratic equations. The solution involves square roots, so how are

they to be found? Cubic equations led to cube roots, more complicated equations to more complicated approximation methods. Arabic mathematicians were innovative in this area. Nasir al-Din al-Tusi, who appears here, is only one of many in a tradition that reaches well beyond Isaac Newton and Joseph Raphson (whose separate contributions are nicely disentangled by the authors).

The fascinating problem of computing values of π gets a chapter here, embracing Chinese methods as well as those of Archimedes, more modern methods using infinite series, and even some of the almost bizarre feats of ingenuity that have given us many million decimal places of π . There is a chapter on the computation of tables, and another on the numerical solution of differential equations in which Carle Runge takes pride of place. We take a trip through approximate methods for finding areas, and tests for prime numbers. Finally, indicative of a recent switch of emphasis, there is a chapter on the theoretical analysis of algorithms in the sense of recursive functions. This discusses when problems can be said to have a solution in any sense that a machine might implement, and leads to the important result that the existence of unsolvable problems can be proved.

Of course, the trichotomy with which I began is an artificial one. The production of an algorithm, in the sense of a systematic solution method, has to be accompanied by an analysis of its range, limitations and proofs of its validity. This is a far from routine task — even though one marvels at the energy of human computers in churning out tables and other numerical results before the electronic computer came along — and the book skilfully shows the intellectual ingenuity that has been expended. Even more important, thinking algorithmically is a distinctive way of approaching mathematics, permeating the entire threefold division and not only operating at the end. Problems about squares are not quadratic equations: they were originally stated without an equation in sight. The answer is not given as a formula that one must fill up with numbers, but as a process. Today we think through problems involving large amounts of data in ways that readily allow for the incorporation of new data. It is not important to calculate more and more of π ; it is most important to use a computer quickly and accurately.

The book is remarkable for its generous quotations from original sources, which allows the discoverers to speak for themselves, and for the wealth of examples that enable us to grapple directly with the material. It has been well edited, and smoothly translated by Chris Weeks. The lonely work of mathematicians shines through and anyone interested in what mathematics really involves will enjoy it.

Jeremy Gray is in the Department of Mathematics, Open University, Milton Keynes MK7 6AA, UK.

Science in culture

Quicker than the eye

“Now you see it. Now you don’t,” claimed the censured Peugeot car advertisements.

Martin Kemp

In the era of the photographic image, blur has become an almost universal way to denote motion faster than the eye can discern. But the actual observation of what came to be called the ‘persistence of vision’ long predated photography. The Islamic philosopher Ibn al-Haytham (Alhazen) described the phenomenon clearly during the Middle Ages, and Leonardo da Vinci in the early sixteenth century drew our attention to the blurred spokes of a spinning wheel, the circle left in the air by a whirling firebrand and the apparent “rods” of falling rain. The motif of blur first dramatically entered painting more than 100 years after Leonardo’s account with the radiant spokes in Diego Velázquez’s *Las Hilanderas* (The Spinners).

We are now so accustomed to the conventions of blurred motion in photography that an image-maker can assume that we will have no difficulty in knowing what a streaky impression signifies. We are also familiar with the reverse convention of frozen motion, stilled by ‘instantaneous’ photography. Famous examples are the coronet splashes, first captured by the physicist Arthur Worthington in 1909 and commercially exploited by Harold Edgerton at Strobe Alley in the middle of the century (see *Nature* 396, 633; 1998). What happens when the apparently contradictory conventions are brought together in a single image?

This combination is precisely what the improbably named advertising agency Euro RSCG Wnek Gosper exploited in the two images devised in their “Now You See It. Now You Don’t” campaign for the Peugeot 206 GTI car. The advertisements were withdrawn after an adverse ruling from the Advertising Standards Authority that they “gave the impression that the car was being driven at high speed” and “could encourage” irresponsibly fast driving. My intention here is not to stir already troubled waters, but to look at the remarkably sophisticated viewing modes that are needed if the images are to operate at full power.

One image, the less perceptually complex, shows a road stretching into the far distance. In the left foreground, a puddle has been splashed



A hummingbird — no car — and speed.

upwards to form a jagged trough frozen in time and space. No car is visible. The perceptual implication is that the car has disappeared in the trice that the splash remains in the air — provided the audience is one that assumes cars drive on the left side of the road.

The second image smears a red blur across the rightmost portion of a road shown parallel to the plane of the picture. On the extreme left, a bird with a rapier-sharp beak is suspended in front of the open calyx of a pink flower. It would be possible to interpret the image simply as an attractive picture of a grand landscape — with the blur of a fast vehicle on the open road. But its full effect depends on our recognizing the bird as a hummingbird and knowing that its wings vibrate so incredibly rapidly when hovering (at 50–80 beats per second) that they would appear blurred in all but a high-speed photographic process. Where has the car gone in that tiny instant?

Taken together, the computer-processed photographic pictures used in the advertisements assume very sophisticated levels of visual knowledge in the target population (readers of the national press). They also show how agencies now expect us to work readily with the remarkably elliptical approach adopted in much of today’s advertising, in which the literal image of the product has been superseded by visual implication and juxtaposition of an often surrealist kind. The viewing habits required are not just those of the photographic era, but also of the age of the digitally manipulated picture. ■
Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.

Where do babies come from?

William Harvey spent a lifetime searching for the earliest moments of life.

R. V. Short

William Harvey (1578–1657) was one of the most distinguished physicians and scientists of the seventeenth century. He was trained by the great anatomist Fabricius in Padua, and, on returning to England, eventually became physician to King Charles I. He was to achieve international acclaim when, in 1628, he published the slim volume *De Motu Cordis et Sanguinis in Animalibus* (On the motion of the heart and the blood in animals). This was the first accurate description of the circulation of the blood, based on acute clinical observations and experiments on live animals.

But Harvey's lifelong obsession was with the far more mysterious subject of reproduction. How was life transmitted from generation to generation? How was it conceived?

Harvey was an Aristotelian at heart, and Aristotle had a 'seed and soil' concept of reproduction, believing that the male provided the seed, which united with the menstrual blood to form an egg. Thus the egg was a product of conception, not something produced by the ovary.

Harvey used an experiment of nature to test Aristotle's views. He chose to study the generation of deer, which are strict seasonal

breeders. His patron, King Charles, hunted deer once a week in his royal forests, parks and chases, and was fascinated by Harvey's studies of generation.

During the rutting season, which starts in late September for red deer and early October for fallow deer, the stags and bucks, with hard-horn antlers and passions inflamed by high levels of testosterone, become too dangerous for the huntsman to dispatch with a thrust of his sword. In Harvey's own words: "Their lust enrages them so, that they will assault or Doggs or Men, when at other times they are shie and timorous, and suffer themselves to be chased and put to flight upon the alarme of the least barking curre that is." So, from September to December, the quarry switched to hind and doe, and Harvey could search for Aristotle's eggs in the lumen of the uterus. Imagine his confusion, therefore, when he could not find any product of conception until early November.

This forced him to conclude that "the foetus doth neither proceed from the seed of male and female emitted in coition, nor yet from any commixture of that seed, (as the Physitians will have it) nor yet out of the Menstruous blood, as Aristotle conceits; and likewise that there is not any thing of the conception necessarily in being,

presently after coition. It is also evident, that all females, in the act of coition, do not essund a seed into the uterus."

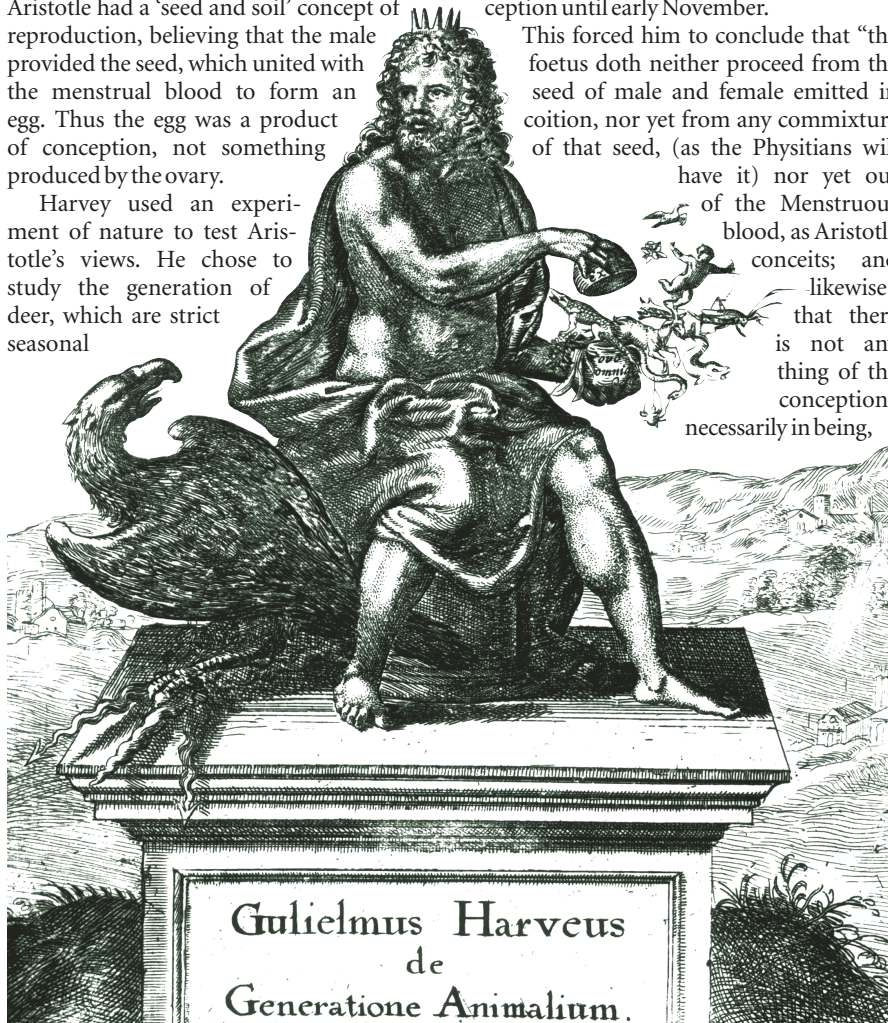
What Harvey failed to appreciate was that although the stags begin to rut in late September, this precedes mating and ovulation by 2–3 weeks. Furthermore, without the aid of a yet-to-be-invented microscope, he would have been unable to see the 0.1-mm-diameter fertilized egg until it had started to elongate, about ten days after fertilization.

And so he was left with a riddle: "Since I plainly see that nothing at all doth remaine in the uterus after coition, whereunto I might ascribe the principle of generation; no more than remaines in the braine after sensation, and experience, whereunto the principle of Art may be reduced; but finding the constitution to be alike in both, I have invented this Fable. Let the learned and ingenious flock of men consider of it; let the supercilious reject it: and for the scoffing ticklish generation, let them laugh their swinge. Because, I say, there is no sensible thing to be found in the uterus, after coition; and yet there is a necessity, that something should be there, which may render the female fruitful."

No doubt it was this bafflement that made Harvey reluctant to publish his findings. After much persuasion, he eventually relented, and *De Generatione Animalium* (On the generation of animals) was published in 1651. By then, Harvey was 73 and nearing the end of his days. Ironically, he chose for the frontispiece a drawing of Zeus holding an Aristotelian egg — from which plant and animal life bursts forth — inscribed with the prophetic words "*Ex ovo omnia*" (Everything comes from an egg).

Harvey would no doubt be mystified by our disapproval of those animal experiments that had enabled him to discover the circulation of the blood, and our condemnation of the hunting of deer with hounds. But a wry smile might have crossed his lips when he read in *Nature* of the cloning of Dolly the sheep, showing that something in the cytoplasm of the enucleated, unfertilized egg has the ability to restore totipotency to a differentiated somatic cell nucleus. He would also be pleased to hear that it is the egg that transmits mitochondrial DNA from generation to generation, and that it is these maternally derived mitochondria that even control the motility and hence the fertility of the male's sperm. *Ex ovo omnia* indeed! ■

R. V. Short is in the Department of Obstetrics and Gynaecology, University of Melbourne, Royal Women's Hospital, 132 Grattan Street, Melbourne, Victoria 3053, Australia.



Divine rite: Zeus brandishes the all-powerful egg in Harvey's *De Generatione Animalium*.

Total internal reflection

The secret of the accidental immortals can now be made known.

Gwyneth Jones

They walk among us. They don't look young — you'd place them around 25 to 30: but the astonishing truth is that they are all (maybe half a million of them world-wide) over 600 years old. They have been talking to journalists, appearing on our screens: they've convinced us that this is no hoax. But why have they decided to leave Earth? That was the question I most wanted to ask, when I was offered the chance to interview our own, local, Thames Valley immortal. Why quit now, just when they don't need to hide anymore?

I met Tamsin in the garden of her house in a quiet Middlesex village: a light-skinned, dark-haired woman of average height, dressed in the dateless human uniform of blue jeans and white T-shirt. She reminded me that I'd agreed not to make a live broadcast. I let her check the output setting on my eye-socket ConjurMac, and we got down to business.

"So," I said (never one to avoid the obvious) "how does it feel to be 650?"

Tamsin laughed.

"How old are you?"

I am 97, and I said so.

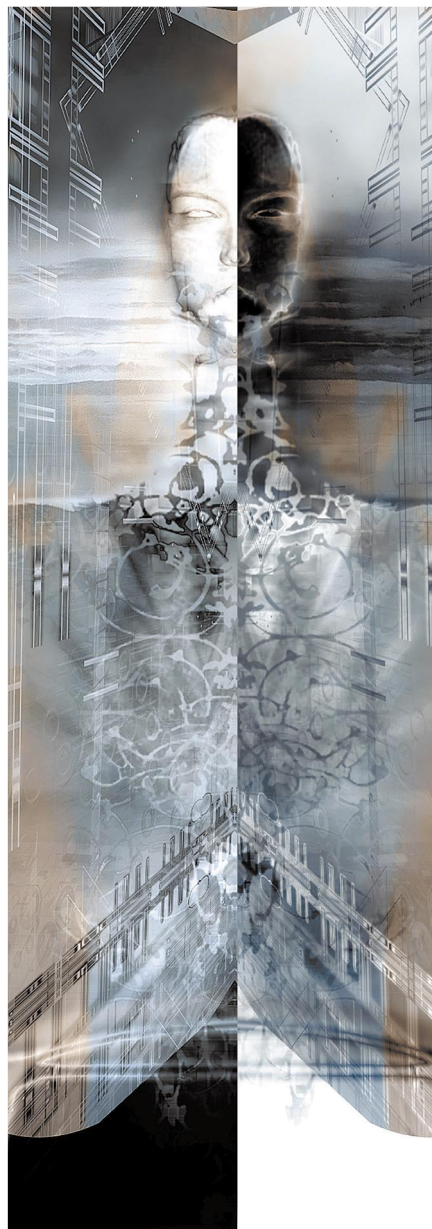
"So why ask me? You'll find out soon enough."

I put it to her that if I survived — and no longevity treatment can *guarantee* survival — it was a long time to wait for an answer.

"When I first took rem," she said. "It was 2039 CE. It doesn't seem long ago at all. Trust me: the years will fly."

Their perception of time is different from ours: I stared, transfixed, at the woman who had lived through the squalor of the Population Pulse, survived five World Wars, kept her impossible secret since the fifteenth century after the Prophet (Praise And Blessings Of Allah Be Upon Him). If I hadn't known, I would never have guessed. She looked so normal.

In the mid-twenty-first century CE, a new drug treatment for memory impairment went into clinical trials. It was meant to strengthen associative memory: in fact it gave patients bursts of recall so intense they were stopped in their tracks, lost in ecstatic re-experience of some childhood joy. Bootlegged rem quickly reached the streets, and



became fashionable. This was in the midst of the Third World War: reckless times. It didn't worry the users that the heart-stopping delight of a rem high was, in a few cases, literally heart stopping.

But the damage mounted. Prodromal schizophrenic symptoms, untreatable depression, vegetative coma — the clinical trials were dropped. All recreationals were legal by then, at least throughout Europe and most of the United States, but rem's reputation made it a poor business prospect. It vanished. The dosage given to the clinical subjects had been too low for the effect to emerge. When those who had taken it habit-

ually realized what was happening to them, they kept quiet. So no one knew, except the immortals themselves.

They were wise to keep quiet. In those days and for a long time afterwards, the Population Pulse was a terrifying force. Longevity research had been abandoned: it was just too sensitive an issue. Rem was not intended to prolong life, only to ameliorate dementia and confusion.

But the Pulse is over, and other things have changed. It's not just the Lagrange colonies and the Moon, and the gruelling labour of love that is transforming Mars. Crucially, in a few months' time the first colony ships will leave Deep Space (our waystation on the brink of the heliopause), and travel, at speeds believed logically impossible when Tamsin was young, to a remote-surveyed, uninhabited, Earth-type planet. This year, 2108 (2688 by the CE count), we are free, at last. The treatments came too late for me. I may live to be 600, but I will get old. For younger generations, there are no known limits.

But rem immortality is still different.

We chatted for two hours, sitting under an apple tree; the murmur of the Fleet (which runs by her garden) in the background. She talked about her son, so long dead; and her husband, killed in a climbing accident last century (immortality is not proof against blunt instruments, lynch mobs or equipment failure). She tried to explain how it feels to have your entire life available to you, so you can be there, again and again, in every moment. How you start with the good moments, then gather courage until you have taken possession of the whole, and your entire existence becomes coherent, like laser light in a perfect optical computer.

"But why are you leaving?" I insisted. "You have so much to teach us."

"It isn't a decision," she said. "It was a process of conversion."

She stood up, smiling, and walked away from me. The air around her shimmered. Next moment, I was alone. ■

Gwyneth Jones's recent books include *Phoenix Café* (Gollancz) and the non-fiction *Deconstructing the Starships* (University of Liverpool Press).

Taiwan's gift to the world

Jared M. Diamond

Study of the giant Austronesian language family tells us a great deal about the history of Pacific peoples and boatbuilding, as well as about Aboriginal Australia.

We humans are defined and fascinated by our languages. Especially intriguing are the 1,200 or so languages of the Austronesian language family, possibly the largest family among the 6,000 languages of the modern world¹. Until the European colonial expansion spread Indo-European languages far and wide after AD 1492, Austronesian was the most widely distributed family, spoken across a realm spanning 26,000 km from Madagascar in the west to Easter Island in the east (Fig. 1).

Austronesian history has been difficult to reconstruct, however, because there are no preserved samples of writing in any Austronesian language until about AD 670, by which time the family's expansion was nearly complete. A reanalysis of Austronesian languages by Robert Blust² strengthens the identification of the first Austronesian waystation, illuminates archaeological findings and the history of boatbuilding, and may help reinterpret the histories of other language families.

Blust's analysis yields an astonishing pattern. Those 1,200 Austronesian languages fall into ten subgroups, of which nine (containing only 26 languages) are spoken only by the non-Chinese aborigines of the island of Taiwan. The tenth subgroup encompasses all Austronesian languages outside Taiwan, from Madagascar to east Polynesia — all 1,174 of them. It is as if the Indo-European language family consisted of 1,174 closely related Slavic languages, spoken from Britain to Sri Lanka, with all nine other Indo-European language groups — Germanic, Celtic, Hittite, Italic and the rest of them — being confined to Ireland. Previous studies had recognized several distinctive Austronesian language groups on Taiwan, but it had not been appreciated that the number was so high.

How do language families differentiate? With time, languages change, and dialects that at first are mutually intelligible gradually become more and more distinct. So it seems that the early diversification of existing Austronesian languages must have taken place long ago, on Taiwan. Eventually, just one group of Taiwanese emigrated to other islands, and their descendants in turn emigrated to still other islands, to become ancestral to all living Austronesian peoples outside Taiwan.

This linguistic evidence for the Austro-



Figure 1 The geographical span of Austronesian languages. This language family encompasses all languages spoken on all Pacific islands from Sumatra in the west to Easter Island in the east, except for the Papuan languages of New Guinea and a few adjacent islands. They are also spoken in Madagascar and in mainland Malaysia. From the work² discussed here, it turns out that of the ten subgroups of Austronesian languages, nine are confined to Taiwan (red circle), and that all Austronesian languages outside Taiwan belong to the tenth subgroup (green), which includes Polynesian languages (dark green; only a few of the hundreds of Polynesian islands are shown here). (Redrawn from ref. 1.)

nesian expansion correlates well with archaeological evidence. Studies of pots, tools and bones have shown that all farming in the Pacific outside New Guinea stems from the colonization of Taiwan by south Chinese farmers by around 4300 BC, followed by their expansion through the Philippines and Indonesia to Polynesia, the Malay peninsula and Madagascar^{1,3}. Of course, pots do not talk, and it can be impossible to guess the languages spoken by the pot-makers. But in the Pacific, identifying the potmakers is easy, because all Polynesian islands were uninhabited until the arrival of people making so-called Lapita pots began at around 1200 BC, and there is no archaeological evidence for arrivals of other peoples after them⁴. Because all traditional languages throughout Polynesia are Austronesian, those first potters must have spoken Austronesian languages.

Especially for those of us interested in boats, the details of Austronesian languages prove as instructive as this main pattern. The contrast between big differences among Taiwanese languages and much more modest differences among extra-Taiwanese languages suggests that there was a 'long pause' between the Austronesian colonization of Taiwan and the Austronesian expansion out of Taiwan. But there is also another contrast, within those extra-Taiwanese languages themselves, between non-Polynesian languages and a discrete sub-subgroup consisting of the closely related Polynesian languages. This suggests that there was a further

long pause, between the first colonization of a bridgehead in Polynesia and the subsequent expansion throughout Polynesia.

Both of these linguistically deduced long pauses are confirmed by archaeological evidence. From this it seems that there was a 1,000-year gap (from about 4300 to 3300 BC) between farmers' colonization of Taiwan and their subsequent colonization of the Philippines, and a further 1,000-year gap (from about 1200 to 200 BC) between the Lapita colonization of west Polynesia and the colonization of east Polynesia¹⁻⁴.

Blust suggests that these two long pauses were due to the time required to develop two leaps in boat technology. Crossing the 375-km seas separating Taiwan from the Philippines would have required much better boats than crossing the mere 140-km strait between mainland China and Taiwan. The ship-building revolution that brought the Philippines and Indonesia within reach may have involved the invention of outrigger canoes. Blust identifies many words in extra-Taiwanese Austronesian languages, but none in the Taiwanese languages, for the component parts of these canoes — which, in historical times, were widespread among Austronesian peoples except for the Taiwanese, who only had bamboo sailing rafts. Similarly, the second ship-building revolution essential to mastery of the open oceans separating the islands of east Polynesia may have been the invention of the Polynesian double-hulled platform sailing canoe, rated

by eighteenth-century European seafarers as superior to contemporary European ocean-going ships.

Blust's study may help us to understand another issue in historical linguistics. The 260 or so Aboriginal Australian languages are usually considered to belong to a single language family⁵. That is surprising, for two reasons. First, people have been living in Australia for at least 50,000 years, ample time for repeated differentiation of language families; and Aboriginal Australian history has been without the homogenizing population movements analogous to the spread of Chinese farmers, whereby one language family could replace all others. Second, Australian languages are similar in their sounds but diverse in their vocabularies, leading linguists to consider them related but to try to explain away their divergent vocabularies.

Blust's work on Taiwanese Austronesian languages suggests that Aboriginal Australia's divergent vocabularies should be taken seriously and attempts made to explain away their convergence in sounds. The diversity in Taiwanese languages was formerly overlooked for several reasons, including their similar sound inventories, for instance the lack of so-called palatal consonants (such as 'z', 'j', 'ch' and 'sh' in English). But Blust points to other cases in which similar sound systems have spread over geographically adjacent language families whose distinctness on other grounds is beyond question. Examples are the sharing of click consonants by South Africa's Zulu and Khoisan languages; the sharing of retroflex consonants (pronounced with the tip of the tongue curled back) among the four otherwise very different language families of the Indian subcontinent; and the shared absence of nasal consonants in languages of North America's Pacific northwest.

This sharing of sounds is expected to develop in an area (such as Aboriginal Australia) where each language is confined to a small tribelet, and where all children grow up multilingual so they can understand and marry members of neighbouring tribelets⁶. For instance, a few months ago, while I was sitting around a campfire in New Guinea with a dozen New Guinea highlanders, each of us volunteered how many languages he spoke. It turned out that every one of the New Guineans spoke between five and ten. Today, I and most other native English-speakers who speak French mangle it with an atrocious accent, and the same is true of most native French-speakers attempting English.

Suppose, however, that an English tribelet and a French tribelet, thrown together with 258 others in an area the size of Australia, were forced to seek marriage partners in other tribelets, and were left in isolation for 50,000 years. At the end of that era, there might still be 260 languages with distinct vocabularies and grammars, but French and

English might now be so similar in their sounds that my wife would no longer blush at my fractured French. The multilingualism of Aboriginal Taiwanese and Australians represented the norm for almost all of human history; we *Nature* readers who grow up in big monolingual nations are an aberration of modern times.

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90095-1751, USA.

e-mail: jdiamond@mednet.ucla.edu

1. Bellwood, P., Fox, J. J. & Tryon, D. *The Austronesians: Historical and Comparative Perspectives* (Australian National University, Canberra, 1995).
2. Blust, R. *Symp. Ser. Inst. Linguist. Acad. Sinica* 1, 31–94 (1999).
3. Bellwood, P. *Prehistory of the Indo-Malaysian Archipelago* (Academic, North Ryde, Australia, 1985).
4. Kirch, P. V. *The Lapita Peoples* (Blackwell, Cambridge, Massachusetts, 1997).
5. Ruhlen, M. *A Guide to the World's Languages*, Vol. 1 (Stanford Univ. Press, 1987).
6. Dixon, R. M. W. *The Rise and Fall of Languages* (Cambridge Univ. Press, 1997).

Display technology

Sidestepping the selection rules

I. D. W. Samuel and A. Beeby

Organic semiconductors that can emit light have developed rapidly over the past decade^{1,2}. These materials offer the prospect of flat, and even flexible, displays that operate at low voltage and emit light, giving excellent contrast and viewing angle. The displays use light-emitting diodes (LEDs) made up of thin layers of organic materials sandwiched between suitable contacts (Fig. 1a). When a voltage is applied to the contacts, charges are injected into the device. Opposite charges can meet up and combine to form an excited state known as

an exciton, of which usually only 25% can go on to emit a photon³.

On page 750 of this issue, Baldo *et al.*⁴ report an ingenious way of getting light emission from the other 75%, thereby potentially improving the efficiency of organic LEDs. This will greatly help to reduce power consumption (which is especially important for portable devices), increase operating lifetime, and increase light output.

Light emission in organic materials commonly occurs when the material absorbs energy and becomes excited, then rapidly

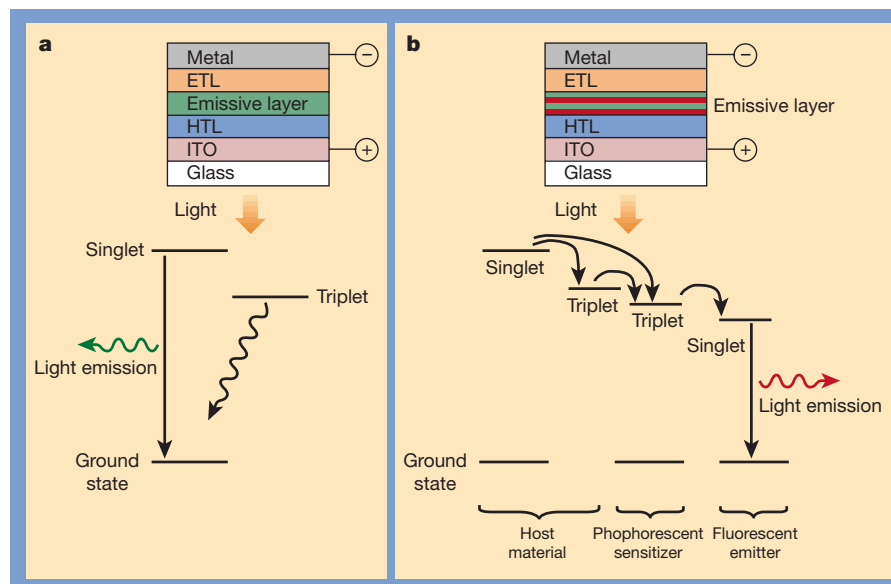


Figure 1 A fourfold increase in the efficiency of an organic light-emitting device. a, A conventional organic light-emitting diode consists of a number of organic layers in between suitable contacts. Opposite charges are injected from each contact and pass through the organic layers and combine to form an exciton in either a 'singlet' or 'triplet' spin configuration. Only the singlet excitons emit light, with the triplets decaying to heat and being wasted. Positive charges move through the hole transport layer (HTL), and negative charges move through the electron transport layer (ETL) to reach the emissive layer where the excitons form. b, In Baldo and colleagues' device⁴, the emissive layer consists of a set of alternating layers, which contain either a phosphorescent material or a fluorescent dye embedded in a host organic material. The result is that both singlet and triplet excitons are transferred from the host material to the phosphorescent sensitizer and on to the fluorescent dye, which then emits light. This avoids the triplet excitons being wasted, greatly increasing the efficiency of light emission. (ITO is a transparent contact, indium tin oxide.)

returns to its ground state by emitting fluorescence. Light emission can also occur by phosphorescence, which gives longer-lived emission. These processes share the same general mechanism but involve different excited states with different electronic configurations. An electron has spin, which can be either up or down, and when two electrons have opposite spins they are said to be spin-paired or in a 'singlet' state. The ground electronic state of most molecules is a singlet state and fluorescence usually involves an excited singlet state relaxing back to the singlet ground state.

Alternatively, if the two electrons of the excited state have the same spin (both up or both down) then it is called a 'triplet' state, and relaxation back to the singlet ground state results in phosphorescence. Transitions of this type are forbidden by the quantum-mechanical selection rules, which in practice means they have low probability and occur more slowly, allowing other processes such as quenching to occur. This generally means that phosphorescent materials emit light with low efficiency.

In an organic LED, the injected charges also have spin, and the way they combine in the device determines the spin configuration of the exciton. This in turn determines whether or not the exciton can emit light efficiently. Unfortunately, only 25% of them become singlet excitons and emit photons; the other 75% become triplet excitons and decay by other pathways. This means that three-quarters of the excitons formed in an organic LED are wasted. So there has been considerable interest in finding ways to obtain light emission from triplet excitons. One approach has been to add phosphorescent compounds to one of the layers in an LED^{5–12}. Triplet excitons can then transfer to these phosphorescent molecules and emit light. However, this method has problems. Relatively few materials phosphoresce efficiently at room temperature, and the long lifetime of the triplet state means that all the phosphorescent sites can become occupied and therefore unavailable. The long lifetime also increases the risk of triplet–triplet interactions that can quench light emission, leading to loss of efficiency, particularly at high intensities^{9,10}.

In Baldo and colleagues' device⁴ both singlet and triplet excitons transfer energy from a host material, first to a phosphorescent molecule and then to a fluorescent molecule, which emits the light. For this to work, the energy of the exciton must be reduced at each step — like a ball rolling down a staircase (Fig. 1b). There are two main mechanisms for energy transfer: the first (known as Dexter transfer) occurs over short distances, effectively requiring contact between the donor and acceptor molecules. An exciton transferred in this way retains its spin configuration (that is, a singlet stays a singlet, and a

triplet stays a triplet). The second, known as Förster transfer, does not require contact, and so can occur over longer distances. Excitons transferred in this way can change spin configuration.

Baldo *et al.*'s aim was to transform both singlet and triplet excitons in the host to the singlet level in the fluorescent dye. They did this using a carefully selected intermediary — a phosphorescent organometallic compound. Both the singlet and triplet excitons transfer their energy to the triplet state of the phosphorescent dye, which in turn can transfer the energy to the excited singlet state of a fluorescent material, which then emits light. Both of these energy transport steps occur by Förster transfer. But competing Dexter transfer could allow a triplet from the phosphorescent material to transfer to a triplet state of the fluorescent dye where it would not emit.

To avoid this the researchers exploited the different ranges of the transfer processes. By placing the phosphorescent and fluorescent molecules in alternating layers of the device, they ensured that the desired Förster energy transfer process generated the excited singlet state of the fluorescent dye. This state decays rapidly, reducing the number of excitons present in the device, and so reducing the chance of quenching by triplet–triplet interactions, or the chance that all of the phosphorescent sites become occupied. An additional advantage of this approach is that it potentially provides a much wider choice of molecules that can be used — they do not have to be phosphorescent to obtain light emission from triplets generated in the device. The increased efficiency will reduce heating within the device and is likely to increase its operating lifetime.

The work is still at an early stage — the efficiencies obtained by Baldo *et al.* with their red LED are not the best that can be achieved, but the potential is there. However, a serious drawback is the need for a number

of downhill energy steps, so that the energy of the emitted light is much lower than that of the positive and negative charges that combine to form the exciton. This means that although the technique is suitable for red devices, it will be difficult to achieve green and even harder to achieve blue emission. It will be interesting to see whether alternative materials offer any improvements.

In the future we may find ways of preventing Dexter transfer by using molecular design rather than alternating layers in the device structure. Although there are benefits for organic LEDs, the advantages of this research for organic lasers could ultimately be much greater. Electrically pumped organic lasers are still some years away, and a serious obstacle is likely to be losses due to re-absorption of the emitted light by triplet excitons. The approach used here has the potential double benefit of harvesting the triplet excitons and reducing their absorption. Sidestepping the selection rules may help realize organic diode lasers as well as producing better LEDs. ■

I. D. W. Samuel is in the Department of Physics and A. Beeby is in the Department of Chemistry, University of Durham, South Road, Durham DH1 3LE, UK.

e-mail: i.d.w.samuel@durham.ac.uk

1. Friend, R. H. *et al.* *Nature* **397**, 121–128 (1999).
2. Sheats, J. R. *et al.* *Science* **273**, 884–888 (1996).
3. Baldo, M. A., Thompson, M. E. & Forrest, S. R. *Phys. Rev. B* **60**, 14422–14428 (1999).
4. Baldo, M. A., Thompson, M. E. & Forrest, S. R. *Nature* **403**, 750–753 (2000).
5. Baldo, M. A. *et al.* *Nature* **395**, 151–154 (1998).
6. Cleave, V., Yahioglu, G., Barny, P. L., Friend, R. H. & Tessler, N. *Adv. Mater.* **11**, 285–288 (1999).
7. Zhang, X. M., Sun, R. G., Zheng, Q. B., Kobayashi, T. & Li, W. L. *Appl. Phys. Lett.* **71**, 2596–2598 (1997).
8. Hoshino, S. & Suzuki, H. *Mol. Cryst. Liq. Cryst. A* **294**, 313–316 (1997).
9. O'Brien, D. F., Baldo, M. A., Thompson, M. E. & Forrest, S. R. *Appl. Phys. Lett.* **74**, 442–444 (1999).
10. Baldo, M. A., Lamansky, S., Burrows, P. E., Thompson, M. E. & Forrest, S. R. *Appl. Phys. Lett.* **75**, 4–6 (1999).
11. McGehee, M. D. *et al.* *Adv. Mater.* **11**, 1349–1354 (1999).
12. Miyamoto, Y., Uekawa, M., Ikeda, H. & Kaifu, K. *J. Lumin.* **81**, 159–164 (1999).

Physiology

Mining the genome for iron

Jerry Kaplan and James P. Kushner

Iron is essential for almost all organisms, but it can also react with oxygen to generate toxic radicals, and so be harmful. Organisms deal with this problem by tightly regulating the concentration of iron in their internal fluids. Humans and other mammals cannot excrete iron. Instead, we control its absorption by the cells lining the gut — the enterocytes — and its export from the enterocytes to the blood.

Although the essential aspects of mammalian iron physiology were established decades ago^{1,2}, the transporters responsible

for absorption of iron into the enterocytes have only recently been identified. On page 776 of this issue, Donovan *et al.*³ now report the identification of the transporter required for exporting iron from the enterocytes into the blood plasma.

Three genes involved in regulating intestinal iron absorption have previously been identified (Fig. 1, overleaf). One of these, *HFE*, is responsible for hereditary haemochromatosis⁴. Mutations in *HFE* result in excessive intestinal iron transport and haemochromatosis, a common disorder

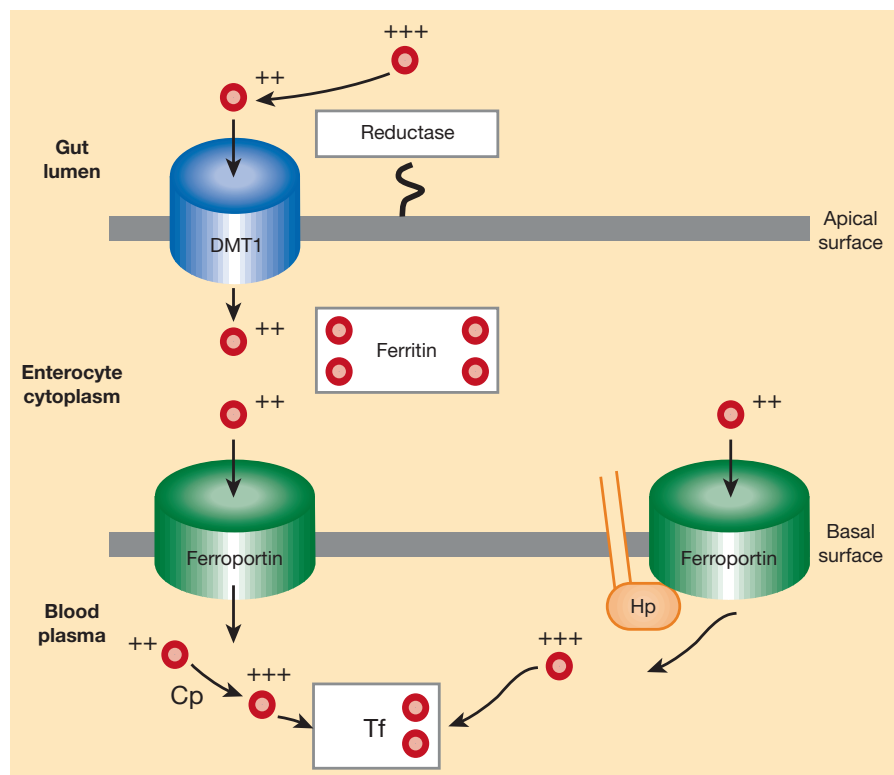


Figure 1 Iron transport across the cells lining the gut (enterocytes). Ferric iron (red dot) in the intestinal lumen is reduced and transported across the apical surface by DMT1. Once inside the cell, iron can either be stored in ferritin or transported across the basolateral membrane by Ferroportin, as shown by Donovan *et al.*³. Ferrireductase activity at the apical surface has been demonstrated, but this activity has not been defined in molecular terms. The species of iron that is exported by Ferroportin is not known, but the need for ferroxidase activity to load iron onto transferrin (Tf) indicates that it is probably Fe^{2+} . Ferroportin1 and Hephaestin (Hp) may work together in the absorptive enterocyte. In other cell types, Ferroportin1 may work with ceruloplasmin (Cp), a soluble ferroxidase, to load iron onto transferrin. Hephaestin-independent iron export may also occur in the enterocyte.

in humans in which iron overload leads to organ damage. Another gene, *DMT1*, encodes the transporter responsible for importing iron from the lumen of the gut into the enterocytes^{5,6}. It was identified in a mouse mutation characterized by iron-limited erythropoiesis (production of red blood cells) and microcytic anaemia (where red blood cells are smaller than normal)⁵.

The protein Hephaestin, the product of the third gene, was identified in mice with the mutation sex-linked anaemia, which also results in microcytic anaemia due to iron-limited erythropoiesis⁷. Hephaestin facilitates the transport of iron from enterocyte to plasma, but is not a membrane transporter. Until now, the transporter that exports iron from the enterocytes into the plasma had not been identified.

Donovan *et al.*³ identified this exporter by cloning a mutant gene in zebrafish, *weissherbst*, that results in iron-limited erythropoiesis. An advantage of using zebrafish for these studies is that they are semi-transparent, so mutants with abnormal blood development can be identified by visible differences in red-blood-cell mass⁸. The authors used mapping studies

to find the part of the zebrafish genome that contained the *weissherbst* gene. They then used DNA sequences spanning that region to probe a zebrafish complementary DNA array to identify expressed sequences encoded by genes in that area. Two of the identified genes were localized to part of human chromosome 2, suggesting that the order of marker genes is conserved among vertebrates.

Another species of fish, the pufferfish, shows conservation of gene-marker order with humans across much of the genome, but pufferfish have much less intragenic and intronic DNA (stretches of DNA found in between and embedded within genes). This results in a very high gene density in any fraction of pufferfish DNA. By analysing a stretch of pufferfish DNA that contained both marker genes, Donovan *et al.* revealed further genes involved in iron regulation. One of them, *ferroportin1*, turned out to be the normal allele of *weissherbst*, and the authors went on to identify homologues in both mouse and human.

They then showed that *ferroportin1* messenger RNA is found in tissues involved in iron transport, namely the zebrafish yolk sac

and intestine, and mouse placenta, intestine and macrophages. Much of the iron that enters plasma comes from the recycling of iron from red blood cells by macrophages, and *ferroportin1* may also be the macrophage iron exporter. The protein was localized to the basolateral surface — the side facing towards the blood — of gut epithelial cells in both mice and zebrafish (Fig. 1), and to the basal surface of the syncytiotrophoblasts, which are specialized cells in mouse placenta. In *weissherbst* mutant fish, iron is not transported across the yolk sac or the gut. Donovan *et al.* showed that Ferroportin1 is an exporter using the *Xenopus* oocyte expression system. Expression of *ferroportin1* in oocytes that also express *DMT1* results in iron export from the cells to transferrin, an iron-binding protein, in the medium.

The iron-uptake transporter on the apical (intestinal) surface of enterocytes, DMT1, co-transporters H^+ and a divalent cation, but nothing is yet known about either the metal selectivity or mechanism of Ferroportin1-mediated iron export. The ability to reconstitute iron export by expression of both DMT1 and Ferroportin1 in *Xenopus* cells should allow a detailed, mechanistic investigation of iron export. This approach will also lend itself to the study of Hephaestin, and will allow researchers to find out how it facilitates iron export.

Finally, the identification of the iron exporter should lead to a better understanding of hereditary haemochromatosis, as the *HFE* gene product ultimately regulates intestinal iron transport. Donovan *et al.* point out that the structure of the *ferroportin1* messenger RNA might indicate that its translation into protein can be regulated by the interaction of an iron-regulatory element with iron-regulatory binding proteins. It remains to be seen whether this is so, and the relative roles of transcriptional and translational regulatory mechanisms need to be clarified.

The identification of Ferroportin1 as the iron exporter in intestine, placenta and, perhaps, macrophages is an exciting advance. Now that the basolateral iron exporter has been identified, as well as the apical uptake transporter, it will allow a detailed analysis of regulated iron absorption, and a mechanistic approach to the malregulation that occurs in hereditary haemochromatosis. ■

Jerry Kaplan and James P. Kushner are in the Departments of Pathology and Medicine, School of Medicine, 50 North Medical Drive, University of Utah, Salt Lake City, Utah 84132, USA.
e-mail: jkushner@crc.med.utah.edu

1. Crosby, W. H. *Semin. Hematol.* **3**, 3299–3313 (1966).
2. Wheby, M. S. *Gastroenterology* **50**, 888–892 (1966).
3. Donovan, A. *et al. Nature* **403**, 776–781 (2000).
4. Feder, J. N. *et al. Nature Genet.* **13**, 399–408 (1996).
5. Fleming, M. D. *et al. Nature Genet.* **16**, 383–386 (1997).
6. Gunshin, H. *et al. Nature* **388**, 482–488 (1997).
7. Vulpe, C. D. *et al. Nature Genet.* **21**, 195–199 (1999).
8. Amatruda, J. F. & Zon, L. I. *Dev. Biol.* **216**, 1–15 (1999).



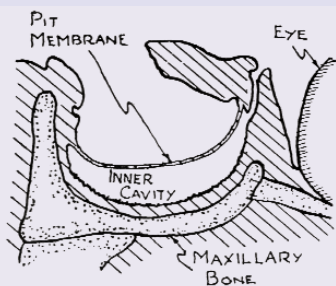
100 YEARS AGO

● The Simplon tunnel is now progressing at the rate of sixteen feet per day. It was begun fourteen months ago, and must be finished in five years and a half from its commencement.

● The proposal of the Council of the Royal Astronomical Society, that the meetings should in future be held at five o'clock, was rejected by the annual meeting on Friday last, and the ordinary meetings will therefore continue to be held at eight o'clock. Taylor's Calendar of meetings of the scientific bodies of London shows that the number of societies which commence their meetings about eight o'clock is more than twice as great as the number of those which meet at four or five o'clock.

● We learn that traffic has just been opened on the Trans-Baikalian section, 700 miles long, of the Siberian railway, as far as Sryétensk. This little town, situated on the Shilka, is in the summer the head of a regular steam-navigation along the Amur, and, with the interruption offered by Lake Baikal, which has still to be crossed on a steamer, Sryétensk can now be reached by rail from Moscow, a distance of about 5000 miles. From *Nature* 15 February 1900.

50 YEARS AGO



Noble and Schmidt have shown that the facial pit of the pit vipers (Crotalinae) is a sensory organ enabling the snakes to detect warm-blooded animals in absolute darkness. Their work indicates that this organ senses the radiation from the warm-blooded animals. The accompanying sketch of the pit from Dr. W. Gardner Lynn's report shows so striking a similarity to a pneumatic radiant-energy detector that one is led to believe that the pit operates in somewhat the same manner; namely, the inner cavity is heated by the radiation and the air in it expands, deflecting the pit membrane. The ultimate branches of the nerves supplying the pit are disturbed in the pit membrane and respond to this deflexion.

From *Nature* 18 February 1950.

Climate change

The hole record

Jonathan T. Overpeck

Will the global warming of recent years continue unabated? A key perspective on that issue comes from what has happened in the past, from palaeoclimatic 'proxy' temperature records that can tell us about conditions before instrument readings became available. On page 756 of this issue, Huang *et al.*¹ provide the latest insights from that source. Their results reinforce the forecast for the twenty-first century that we have heard before: continued warming ahead. But they also provide unsettling indications that human alteration of the climate system over the past century will make the reliable prediction of climate change an even tougher business than expected.

What the authors have done is to compile a large database of underground temperature measurements from 616 previously drilled boreholes on six continents. Because the measured temperature change with depth in each borehole is a function of time-dependent heating from both below (the interior of the Earth) and above (the atmosphere), it was possible to couple the pattern of downhole temperature with the theory of heat conduction to obtain a time history of ground surface temperature for each borehole. The inclusion of many boreholes in their analysis allowed Huang *et al.* to filter the climatic signal from non-climatic noise, and thus obtain convincing new evidence that the Earth is warming above natural levels.

This study is notable for several reasons. First, the authors' borehole source of proxy temperature information is independent of those used in previous palaeoclimatic reconstructions (for example tree rings, ice cores, corals and sediments). Second, they provide a broader spatial coverage and improved density of palaeotemperature records on land than earlier studies (see the map on page 756 of this issue). In both the Southern and Northern Hemispheres, the borehole records match the instrumental records of the past century, and confirm that the twentieth century was the warmest — by about 0.5 °C — of the past 500 years. Moreover, each of six continents investigated has warmed faster during the 1900s than during any of the previous four centuries.

Huang and colleagues' analysis is the latest of several^{2–7} to indicate that late-twentieth-century warming is without precedent in the past 400 to 1,000 years. We do not know of any combination of natural mechanisms that can explain this phenomenon. Variation in solar output, or episodic reduction in the amount of sunlight reaching the



Figure 1 Kiet Siel, an abandoned settlement in Arizona. Although drought was mostly a seasonal to interannual phenomenon during the twentieth century, the palaeoclimatic record is replete with examples (as yet unexplained) of episodes of drought that lasted several decades and had a severe impact on human communities. A classic case is the drought that occurred in the late thirteenth century in the southwest United States. It lasted for over 20 years and led to the widespread abandonment of settlements in the area^{12,13}. Huang *et al.*¹ highlight the possibility that pre-industrial natural temperature variability was also larger than generally estimated. (Picture courtesy of J. Dean.)

Earth's surface because of volcanic activity, seem to account for many features of the pre-industrial portion of the temperature record. But such mechanisms can explain only a fraction of the total warming that took place in the twentieth century^{2,3,7–10}. So we are left with the likelihood that human-induced global warming is under way, caused by emissions of 'greenhouse gases' such as carbon dioxide, and that the next century is going to see even greater warming¹¹.

The question is, how much? Policy-makers, who must look ahead to the consequences of warming and ways of tackling them, need answers as a basis for their decisions. But those answers have to pertain not just to how much warming will occur 50 or 100 years from now, but also to how this warming will affect climate variability, including climatic extremes. We know from the past that natural interannual to century-scale variability is in itself large enough to be a concern for human economies and welfare (Fig. 1). Adapting to future climates will thus require a more complete understanding

of natural variability as well as of human-induced changes.

The results of Huang *et al.*¹ show yet again that the twentieth-century record of climate variability is too short and cloaked with human-induced influences to provide a clear indication of natural climate variability. More importantly, Huang *et al.* go beyond the canonical debate on recent warming to suggest that earlier studies may have underestimated the full amplitude of natural decade-to century-scale climate variability. This is not a complete surprise. We suspected, for example, that tree-ring records of temperature change (used in most previous analyses) may not always preserve all of the long-term variability that influenced past tree growth⁶. Other proxy sources of palaeoclimatic information may have site- or proxy-specific biases as well⁴. What are required are larger, multi-proxy networks of palaeoclimatic records that can complement the denser coverage of recent instrumental stations. Only then will we have the fully reliable estimates

of natural variability that we need to help us adapt to global climate change. ■

Jonathan T. Overpeck is at the Institute for the Study of Planet Earth, and the Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA.

e-mail: jto@u.arizona.edu

- Huang, S., Pollack, H. N. & Shen, P.-T. *Nature* **403**, 756–758 (2000).
- Overpeck, J. T. *et al.* *Science* **278**, 1251–1256 (1997).
- Mann, M. E., Bradley, R. S. & Hughes, M. K. *Nature* **392**, 779–787 (1998).
- Jones, P. D., Briffa, K. R., Barnett, T. P. & Tett, S. F. B. *Holocene* **8**, 455–471 (1998).
- Mann, M. E., Bradley, R. S. & Hughes, M. K. *Geophys. Res. Lett.* **26**, 759–762 (1999).
- Briffa, K. R. & Osborn, T. J. *Science* **284**, 926–927 (1999).
- Mann, M. E. *et al.* *Earth Interactions* (in the press).
- Briffa, K. R., Jones, P. D., Schweingruber, F. H. & Osborn, T. J. *Nature* **393**, 450–455 (1998).
- Crowley, T. J. & Kim, K. Y. *Geophys. Res. Lett.* **26**, 1901–1904 (1999).
- Free, M. & Robock, A. J. *Geophys. Res.* **104**, 19057–19070 (1999).
- Houghton, J. T. *et al.* (eds) *Climate Change 1995* (Cambridge Univ. Press, 1996).
- Dean, J. S. *Am. Antiquity* **50**, 537–554 (1985).
- Dean, J. S. *Clim. Change* **26**, 225–241 (1994).

Reproductive biology

A provirus put to work

Jonathan P. Stoye and John M. Coffin

The genetic information of retroviruses consists of single-stranded RNA, and an essential step in retroviral replication involves the integration of a reverse-transcribed form of that RNA, as double-stranded DNA, into the genome of the infected cell to form a provirus. In this way, retroviruses can colonize the host germ line and be inherited as endogenous retroviruses (ERVs)¹. The upshot, over evolutionary time, is the presence of thousands of proviruses in the human genome.

Retroviruses are usually associated with harmful consequences. But they, or their gene products, can also be beneficial, particularly in preventing infection by other retroviruses. On page 785 of this issue², Mi *et al.* provide evidence that one particular ERV may be essential in human physiology in quite a different way — for the formation of the placenta.

During pregnancy the developing fetus must be both nourished and protected. The placenta has a key part in these processes. In most species, a so-called syncytiotrophoblast layer is present on the outside of the placenta at the fetal–maternal interface. It is a continuous structure, one cell deep, formed by fusion of the constituent trophoblast cells (Fig. 1). The layer has pronounced microvillar surfaces, which presumably facilitate exchange of nutrients and waste products between mother and fetus.

It has been known for some time that

ERVs are expressed in the placenta, leading to speculation as to their involvement in placental formation or function³. For example,

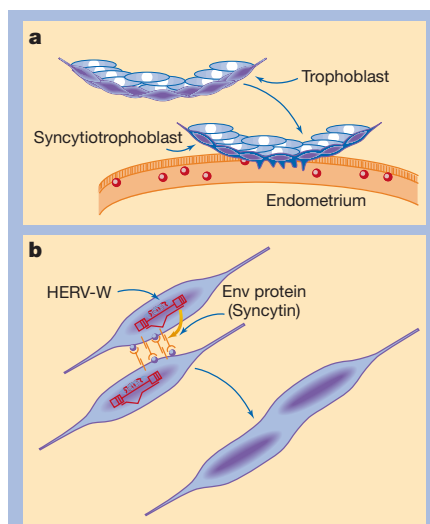


Figure 1 Possible functions of the endogenous retrovirus HERV-W in human placental development, prompted by the paper of Mi *et al.*². a, Fusion of the trophoblast cell layer into the continuous multinucleate syncytiotrophoblast is associated with implantation of the embryo. b, Expression of the HERV-W Env (envelope) protein mediates trophoblast cell fusion, probably by a mechanism identical to that by which the virus enters the host cell.

the viral envelope protein Env (produced by the viral *env* gene) has properties that enable the virus to bind a specific receptor on a target cell, fuse with that cell, and so mediate viral entry. The same properties could mediate trophoblast fusion⁴. Further, an immunosuppressive domain present in the viral transmembrane protein might help prevent maternal immune responses against the developing embryo⁵.

Identification of ERVs has frequently involved serendipity¹. The study by Mi *et al.*² is no exception. During a screen to isolate secreted proteins, the authors identified a complementary DNA predicted to encode a protein with homology to several other retroviral Envs. Strikingly, the protein, subsequently dubbed syncytin, is identical to that encoded by the envelope gene of the human endogenous retrovirus HERV-W⁶. Syncytin is expressed at high levels in the placenta, specifically in the syncytiotrophoblast layer, and at lower levels in the testis, but nowhere else. The provirus that appears to encode syncytin has undergone inactivating mutations in the other viral genes, consistent with the idea that *env* function has been selectively preserved⁶.

Mi *et al.* sought a direct test for how syncytin might be involved in forming the syncytiotrophoblast layer. Introduction of syncytin into culture systems led to cell fusion and syncytia formation; fusion could be blocked by antibodies directed against syncytin. Further studies using an inducible model for trophoblast fusion showed that fusion was associated with an increase in syncytin gene transcription and could also be inhibited by anti-syncytin antibody treatment. Interestingly, syncytin-mediated fusion appears to be rather promiscuous, with no apparent need for a specific receptor. This result is in potential conflict with indications that HERV-W binds the type D mammalian retrovirus receptor⁷ (as well as with the apparent selectivity of trophoblast fusion), though different assay systems may explain this discrepancy.

The new data² are consistent with syncytin mediation of trophoblast fusion to form the syncytiotrophoblast layer. But they do not constitute conclusive evidence. How might one go about obtaining such evidence? The HERV-W family of ERVs is restricted to primates, which rules out direct gene knock-out experiments. An alternative approach, taken in studies with ERV-3, is to look for naturally occurring genetic polymorphisms that might result in gene inactivation. About 1% of the human population is homozygous for a mutation of ERV-3 that would cause premature termination of envelope-protein translation, effectively meaning that ERV-3 cannot be essential for placenta formation⁸. Analogous studies with syncytin will be more complicated because there are 10–15 copies of HERV-W

in the human genome. Such studies can be done, however, and could be extremely rewarding.

Adaptation of retroviral functions to serve the needs of the host requires, at a minimum, that the provirus is expressed at the right time and in the right place, and that the co-opted function meshes with host-cell systems (such as receptors). It is likely, although it has not yet been shown, that syncytin expression is directed by the normal promoter in the proviral long terminal repeat. Whether the properties of placenta-specific expression and interaction with the correct receptor were inherent in the virus ancestor, or were acquired subsequent to integration, remains to be seen. In any case, even small changes have been found to dramatically alter both properties in other retroviral models.

The possible involvement of ERV in placental physiology has led to speculation that retroviruses might have had a role in the evolution of all placental mammals^{3,5}. In its simplest form, such a hypothesis would require the existence of a single conserved provirus or proviral gene at a defined location in all species. This does not appear to be the case — individual ERV families seem to be relatively restricted in their species distribution. For example, HERV-Ws are present only in Old World primates. So syncytin cannot be responsible for trophoblast fusion in, say, mice. It is conceivable that different families of ERV might have the same function, but this is unlikely if the function concerned is an essential one.

More probably, specific retroviral integrants have niche functions that have conferred minor evolutionary benefits on different species at different times. For example, mice have adapted several ERV gene products to generate genes providing resistance to retroviral infection⁹, and a retroviral insertion upstream of a duplicated copy of the pancreatic amylase gene may be responsible for the ability of certain primates, including humans, to taste starchy material as sweet¹⁰. The work of Mi *et al.* suggests that the list of niche roles may be extended to include functions in placental development. Indeed, it is tempting to speculate that the diversity of placental structures seen in nature might in part reflect the use of different ERV products to facilitate placental function.

Jonathan P. Stoye is in the Division of Virology, National Institute for Medical Research, Medical Research Council, The Ridgeway, Mill Hill, London NW7 1AA, UK.

e-mail: jstoye@nimr.mrc.ac.uk

John M. Coffin is in the Department of Molecular Biology and Microbiology, Tufts University School of Medicine, 136 Harrison Avenue, Boston, Massachusetts 02111, USA.

e-mail: jcoffin_par@opal.tufts.edu

1. Boeke, J. D. & Stoye, J. P. in *Retroviruses* (eds Coffin, J. M., Hughes, S. H. & Varmus, H. E.) 343–435 (Cold Spring Harbor Laboratory Press, New York, 1997).

2. Mi, S. *et al.* *Nature* **403**, 785–789 (2000).
3. Harris, J. R. *BioEssays* **20**, 307–316 (1998).
4. Boyd, M. T., Bax, C. M., Bax, B. E., Bloxam, D. L. & Weiss, R. A. *Virology* **196**, 905–909 (1993).
5. Villareal, L. P. *J. Virol.* **71**, 859–865 (1997).
6. Blond, J. L. *et al.* *J. Virol.* **73**, 1175–1185 (1999).

7. Blond, J.-L. *et al.* *J. Virol.* (in the press).
8. de Parseval, N. & Heidmann, T. *J. Virol.* **72**, 3442–3445 (1998).
9. Best, S., Le Tissier, P. R. & Stoye, J. P. *Trends Microbiol.* **4**, 313–318 (1997).
10. Meisler, M. H. & Ting, C. N. *Crit. Rev. Oral Biol. Med.* **4**, 503–509 (1993).

Superconductivity

From obscurity to impurity

A. V. Balatsky

During the past decade we have seen impressive applications of scanning tunnelling microscopy to the study of different materials, including superconductors^{1,2}. Since the discovery of superconductivity at relatively high temperatures (up to 160 K) in compounds containing copper oxide layers, physicists have been trying to understand the basic mechanism, and it ranks as one of the most fundamental questions in condensed-matter physics. Scanning tunnelling microscopy (STM) promises a new approach to the study of high-temperature superconductors. The report by Pan *et al.*³ (on page 746 of this issue) presents STM imaging and tunnelling spectroscopy of the electronic states found around impurity atoms, which have been swapped for copper atoms in the copper oxide plane of a high-temperature superconductor.

The experiment by Pan *et al.* is the latest example of the importance of STM, a technique that is sure to assist in the ultimate solution of the high-temperature superconductor puzzle. When the metallic tip of an STM is scanned across a conducting surface, electrons quantum-mechanically tunnel between the tip and the surface. Tunnelling spectra record changes in the tip current as it passes over the electronic states at the surface. One of the first successes of STM spectroscopy was mapping electronic states in a

superconductor. Superconductivity develops when the electrons bind together in pairs, forming a so-called condensate. This usually occurs in metals at extremely low temperatures of a few kelvin, when the electrons in the condensate begin to move without resistance. In a high-temperature superconductor, electrons are confined within the conducting layers of copper oxide, and the unusual properties of these compounds appear to be related to the correlated motions of the electrons within these layers.

Why are impurities so important to the study of high-temperature superconductivity? We know that in its pure state a high-temperature superconductor is complicated — why make it more so? The idea is that impurities interfere with the mechanism creating the superconductivity, destroying it in the vicinity of the impurity and creating a localized state, which STM can probe on an atomic scale. By discovering how the superconductivity is destroyed, we hope to better understand the mechanism behind it. This approach is similar to a child taking a toy apart to see how it works. Pan *et al.*³ have succeeded in obtaining clear STM images and electron-tunnelling spectra of the impurity states induced by zinc impurity atoms in their copper oxide superconductor.

We know that high-temperature superconductors have a characteristic energy gap,

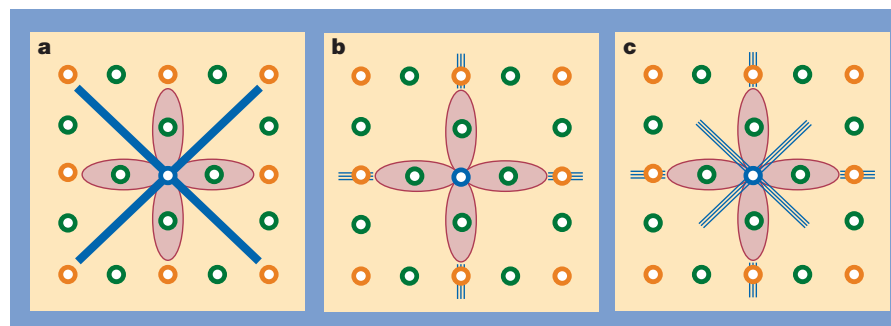


Figure 1 Energy gap in a high-temperature superconductor. a, Effect of an impurity atom (zinc, blue) on the electronic states of the copper oxide layer of a superconductor (copper atoms are orange and oxygen atoms are green). The superconducting energy gap is pink and the predicted impurity state is a blue cross. b, The interlayer tunnelling conductance. These materials are known to be stacked in layers with relatively small tunnelling between layers. The interlayer tunnelling is strongest along Cu–O bonds in the lattice. c, The net tunnelling signal recorded by Pan *et al.*³. Overlapping the impurity intensity and interlayer tunnelling intensity produces a mixed signal that reduces the size of the cross and emphasizes the horizontal and vertical intensities.

unlike anything seen in ordinary metals. The energy gap is the minimum energy required to produce the first electronic excitation in the superconductor. This gap is largest along the copper oxide bonds and becomes zero only along the diagonals of the square CuO_2 lattice (Fig. 1a). By locally interfering with the mechanism for superconductivity in the CuO_2 plane, the zinc atom should produce an impurity state that looks to an STM like a high-intensity cross rotated at 45° to the energy gap (Fig. 1a). The simplest analogy would be a cup nearly filled with water. To escape, the water must leap over the rim of the cup, which would require a certain amount of energy. If four notches are carved in the rim (to simulate the zeroes of the energy gap in the high-temperature superconductor), the water will leak out of the notches. Similarly, electrons in the impurity state 'leak out' at the zeroes of the energy gap. This effect was predicted several years ago⁴⁻⁶, but Pan *et al.*³ are the first to find one of these cross-like impurity states.

Surprisingly, when the CuO_2 layer is not at the surface, the impurity 'cross' seen by the STM at greater distances from the impurity seems to be more aligned with the gap. One possible explanation for this observation is that, because the impurity is not on the surface, the measured signal is a mixture of the impurity state and the interlayer tunnelling conductance (Fig. 1b). Tunnelling electrons from the impurity state must reach the STM tip above the surface of the material. To get to the outside, the electron has to go through interlayer tunnelling that is actually larger (by a factor of ten or more) in the horizontal and vertical directions (along the copper oxide bonds in the lattice). The net signal is shown in Fig. 1c. More experiments are needed to see how relevant this idea might be.

One can imagine other STM experiments using impurity atoms in high-temperature superconductors in the future. An interesting and largely unexplored aspect is the effect of magnetic impurities on the superconducting state. Pan *et al.*³ studied the effects of a zinc impurity, whose influence is primarily due to its net charge. Other impurities, such as nickel, have a magnetic spin — akin to a tiny magnet inside the atom — when substituted into a superconductor. Such magnetic impurities have previously been studied by STM in the context of conventional superconductors⁷. In high-temperature superconductors their effects are likely to be significant.

In these materials with no conducting electrons, the ground state is an antiferromagnet, such that neighbouring copper atoms have alternating spins aligned in opposite directions, say up-down-up-down as we move along copper-oxygen-copper bonds. The antiferromagnetic nature of the copper oxide layer is a crucial ingredient

in the mechanism producing high-temperature superconductivity. STM studies of magnetic impurities will help us to understand better the interplay between magnetism and high-temperature superconductivity at the atomic scale. For example, we might observe the shape and energy characteristics of the superconducting state near a magnetic impurity. Magnetic correlations near an impurity could restore the alternating copper spins and produce a 'staggering effect', whereby the positive and negative voltage peaks in the tunnelling data would appear on alternating atoms, corresponding to the sublattice structure of the antiferromagnetic ground state.

We might also expect to see the original superconducting state not just suppressed by magnetic impurities, but also changed near an impurity. For example, the impurity may distort the original perfect four-leaf gap

structure (Fig. 1a), generating additional features in the gap. A colleague and I have previously proposed that patches of complex gap components are created as a result of condensate interactions with magnetic impurities^{8,9}. These and other local effects are perfectly suited for STM measurements. ■

A. V. Balatsky is at the Los Alamos National Laboratory, Theory Division, MS B 262, Los Alamos, New Mexico 87545, USA.

e-mail: avb@lanl.gov

1. Yazdani, A. & Lieber, C. M. *Nature* **401**, 227–230 (1999).
2. Atkinson, W. A. & MacDonald, A. H. *Science* **285**, 57–58 (1999).
3. Pan, S. H. *et al.* *Nature* **403**, 746–750 (2000).
4. Balatsky, A. V., Salkola, M. & Rosengren, A. *Phys. Rev. B* **51**, 15547–15551 (1995).
5. Stamp, P. J. *Magn. Mater.* **63/64**, 429 (1987).
6. Byers, J. M., Flatté, J. M. & Scalapino, D. J. *Phys. Rev. Lett.* **71**, 3363–3366 (1993).
7. Yazdani, A., Jones, B. A., Lutz, C. P., Crommie, M. F. & Eigler, D. M. *Science* **275**, 1767–1770 (1997).
8. Balatsky, A. V. *Phys. Rev. Lett.* **80**, 1972–1975 (1998).
9. Balatsky, A. V. & Movshovich, R. *Physica B* **259–261**, 446 (1999).

Ecology

The complexity of co-dependency

Peter J. Morin

It is generally thought that ecosystems containing a greater diversity of species are more likely to thrive than those with fewer species. This has led to efforts to conserve or restore biodiversity in natural ecosystems. Diversity at the level of producers (such as photosynthetic algae) and decomposers (bacteria or fungi), in particular, may influence ecosystem productivity. On page 762 of this issue, Naeem *et al.*¹ report an elegant experimental study of the productivity — the amount of biomass generated — of a simple aquatic ecosystem. They find that productivity depends on the diversity of both producers and decomposers, but that the relationship between the two is not straightforward.

Naeem and colleagues' experiment involved independent manipulations of the number of algal and bacterial species in laboratory microcosms. The results were surprisingly complex, and do not support a simple relationship between diversity and productivity. Systems containing a single algal species and many bacterial species displayed a low combined productivity of algae and bacteria, and much of the total biomass consisted of bacteria. In contrast, systems with moderate to high algal diversity tended to have a higher combined productivity characterized by high algal biomass and relatively low bacterial biomass, regardless of bacterial diversity.

Naeem *et al.* also measured the ability of the bacteria in their microcosms to use different organic compounds as carbon sources. They found that bacteria from systems with

high productivity could use a greater variety of carbon sources than those from systems with lower productivity. Interestingly, the ability of bacteria from a microcosm to use a variety of carbon sources did not depend solely on the number of bacterial species present, but also increased with increasing algal diversity. This may mean that ecosystem productivity is linked to the cycling of carbon or other nutrients.

The study also confirms a result previously described only for terrestrial ecosystems²⁻⁴. Ecosystems containing more producer species tend to be more productive, but the pattern varies greatly with the number of bacterial species added to the system. Although they extend the generality of positive relations between diversity and ecosystem processes in terrestrial systems to the aquatic realm, Naeem *et al.* point out that such patterns may depend on whether researchers manipulate only producers (as is often the case) or other groups crucial to ecosystem function, such as predators. Most studies have manipulated the diversity of only producers^{2,3} or mycorrhizal fungi⁵, or of entire food webs⁴, which makes it difficult to see whether the positive effects of diversity on ecosystem properties depend on other unspecified features of species composition.

Complex interactions between producers and decomposers potentially occur in all ecosystems, with the possible exception of systems such as deep-sea hydrothermal vent communities that are not directly powered by the capture of sunlight. Recent models have predicted the complexity of

interactions between producers and decomposers^{6,7}. Positive interactions arise because photosynthetic producers create the organic compounds consumed by decomposers, whereas decomposers remineralize carbon and other nutrients required by producers.

Simpler models treat producers and decomposers as single species, an assumption that now seems suspect given Naeem and colleagues' results. Models that incorporate competition among species within nutrient cycles could better explain the complexity of interactions between producers and decomposers⁷. For instance, low productivity in systems with many bacterial species and few algae could be a consequence of the greater likelihood that those systems contain a bacterial species that tends to outcompete the algae for some essential nutrient. Greater productivity of systems containing a greater diversity of both algae and bacteria would be predicted by the increased likelihood that each group (producer or decomposer) would contain a species that is highly efficient in using resources⁷.

Although aquatic algae and bacteria may compete for limiting nutrients, there is ample evidence that they provide mutual benefits. Algae leak the sugars produced by photosynthesis, especially when nutrient-limited, providing a ready carbon source for bacteria. The total abundances of algae and bacteria are positively correlated in lakes⁸, so it seems that bacteria enjoy some sort of positive association with algae even though they may outcompete them for nitrogen and phosphorus.

In turn, bacteria may provide benefits for some algae beyond the remineralization of macronutrients. Not all producers can synthesize all of the compounds needed for cellular metabolism. Some, particularly photosynthetic flagellates such as *Euglena*, have an absolute dietary requirement for vitamin B₁₂ (ref. 9), which can be synthesized by bacteria. Although *Euglena* is only one of the several algal species used by Naeem *et al.*, it and other species may benefit in similar ways from micronutrients synthesized by bacteria. Another possibility is that some photosynthetic flagellates (*Euglena* again) are mixotrophic — that is, able both to photosynthesize and to consume other organisms, including bacteria. If so, this might explain why, in Naeem and colleagues' experiments, algae sometimes benefited from increased bacterial diversity, whereas bacteria did not appear to benefit from increased algal diversity.

Naeem *et al.* provide a compelling example of the value of laboratory microcosms in unravelling the complex relations between ecosystem diversity and function. Their manipulation of algal and bacterial diversity would be impossible under field conditions. Microcosms also provide opportunities to

test the mechanisms that underlie the positive effects of bacteria on algae. If supplementing microcosms with limiting micronutrients, such as vitamin B₁₂, eliminates the positive effect of bacterial diversity on algal productivity, then a real mutualistic association between algae and bacteria would seem likely. If the positive influence of bacteria is limited to mixotrophic algal species, then a different kind of mechanism involving greater production of mixotrophic consumers on a greater diversity of bacteria prey is possible. Regardless of the mechanism, it is clear that the productivity of aquatic systems depends on more

than the diversity of primary producers. ■

Peter J. Morin is in the Department of Ecology, Evolution and Natural Resources, 14 College Farm Road, Cook College, Rutgers University, New Brunswick, New Jersey 08901, USA.

e-mail: pjmorin@rci.rutgers.edu

1. Naeem, S., Hahn, D. & Schuurman, G. *Nature* **403**, 762–764 (2000).
2. Tilman, D., Wedin, D. & Knops, J. *Nature* **379**, 718–720 (1996).
3. Hector, A. *et al.* *Science* **286**, 1123–1127 (1999).
4. Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. N. *Nature* **368**, 734–737 (1994).
5. van der Heijden, M. G. A. *et al.* *Nature* **396**, 69–72 (1998).
6. Harte, J. & Kinzig, A. *Am. Nat.* **141**, 829–846 (1993).
7. Loreau, M. *Proc. R. Soc. Lond. B* **265**, 33–38 (1998).
8. Currie, D. J. *Limnol. Oceanogr.* **35**, 1437–1455 (1990).
9. Sleigh, M. *Protozoa and Other Protists* (Arnold, London, 1989).

Mathematics

A gathering of groups

Ian Stewart

The branch of algebra known as group theory has its origins in several major areas of nineteenth-century mathematics. In abstract terms, a group is just a set of elements that can be combined together in pairs to yield another element of the set, in a manner that obeys three algebraic laws: the existence of an identity element, the existence of an inverse for any given element, and the associative law $(ab)c = a(bc)$. The elements may be operations, such as the rotations of a cube (Fig. 1). More concretely, group theory is the mathematics of symmetry (and much more), and symmetry is fundamental to many areas of science, notably quantum mechanics.

The number of elements in a group — its 'order' — can be finite or infinite. The finite groups have long intrigued algebraists, because they formalize the concept of 'symmetry' and have applications ranging from the insolubility of the quintic equation to crystallography. Today a huge amount is known about them, including a complete classification of the simplest finite groups — the basic 'building blocks' for all groups. But, as any fond parent is aware, there is a big difference between knowing about individual building blocks and knowing everything that can be built with them.

Hans Besche and Bettina Eick have now applied methods of computer algebra to count and construct all finite groups with 1,000 elements or less — that is, up to order 1,000. Most of this work is reported in the *Journal of Symbolic Computation* (**17**, 405–513; 1999), except for two much harder cases (512 and 768), which are discussed in work currently in press (some jointly with E. A. O'Brien).

Their results illustrate the striking irregularity of the group-theory landscape. For example, there is exactly one finite group of

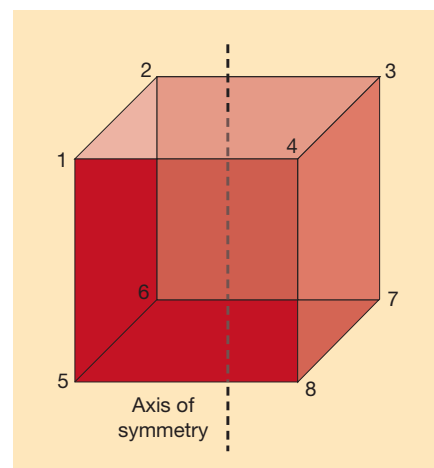


Figure 1 The symmetries of a cube are best described by a group. For example, a rotation of 90° about the axis of symmetry shown is given by a specific permutation of the 8 vertices. A rotation about another axis through the vertex 1 and the opposite vertex 7 is given by a specific permutation that holds the integer 7 fixed. Combinations of these give a total of 24 rotations (including the identity rotation). Apart from these rotations, there are also mirror-image reflections about a plane through the midpoints of the edges. Combining these reflections with the 24 rotations gives the full group of 48 symmetries of the cube.

order 895, but there are 19,349 of order 896 and two of order 897. This irregularity comes about because the structure of a group is to a great extent determined by the factorization of its order into primes (a prime number is divisible only by 1 and itself). For example, there is exactly one group of any prime order. But there are also non-prime orders with only one group — such as 895 — and the smallest case of this type is order 15. The more prime factors,

the more cases the would-be classifier has to analyse, so it is not surprising that the worst cases of all are orders that form powers of two. The numbers of groups of orders 2, 4, 8, 16, 32, 64, 128 and 256 are, respectively, 1, 2, 5, 14, 51, 267, 2,328 and 56,092 — so it is not surprising that order 512 is omitted from the paper mentioned above. The other omission, 768, is equal to 3×256 , again with an unusually large number of prime factors.

Besche and Eick's work relies heavily on a number of general conceptual results proved by group theorists over the past century, in which groups of large order are in one way or another decomposed into bits and pieces that form groups of smaller order. As a rough analogy, any integer can be decomposed into a product of primes, but in group theory the building blocks are more complex than prime numbers, and the ways of combining them are more varied than mere multiplication.

Besche and Eick use three basic methods. Two of these — the Frattini method and the cyclic-extensions method — apply to so-called soluble groups, which can be built up from components of prime order. The third — the upward-extension method — deals with groups that are not soluble. Incidentally, soluble groups bear that name because in Galois group theory they correspond to equations that can be solved by formulae involving nothing worse than n th roots; by a happy twist of fate, soluble groups are far more common than insoluble ones. Mathematics is seldom that obliging, and even

here there is a price to pay: it is more difficult to list all the possibilities for soluble groups than for insoluble ones.

Calculations of this kind would be absolutely horrendous if performed by hand (though this has never stopped mathematicians in the past). Instead, they are carried out using a computer algebra package known as GAP (Groups, Algorithms and Programming) that was specifically designed for group-theory computations. A catalogue of the relevant groups is built into recent releases of GAP, and also into its competitor MAGMA. It is not sensible to perform such classifications using brute force: for all their power, today's computers have a habit of wilting when faced with relatively unambitious questions in pure mathematics.

So sophisticated algebraic techniques, based on the deep modern understanding of how to pull a group to bits and reconstitute it from the pieces, must be brought to bear in order to make the computations tractable. With these techniques we can hope to gain a much more detailed understanding of the possible structures for finite groups, with potential applications to all areas of science where symmetry is important. These range from quantum mechanics to pattern formation in fluids, and include surprising areas such as animal locomotion and evolutionary speciation. ■

Ian Stewart is in the Department of Mathematics, University of Warwick, Coventry CV4 7AL, UK.
e-mail: ins@maths.warwick.ac.uk

itself cause nearby cells to become ventral neurons. This action is mediated by the signalling molecule Sonic hedgehog (Shh), which is expressed first by the notochord and then by floor-plate cells. A mechanism for dorsal specification has remained more elusive, leading to the hypothesis that dorsal fates might be a default state in the absence of ventral induction. The finding that the non-neural epidermal ectoderm, adjacent to the neural plate, is responsible for the induction of neural crest and dorsal phenotypes in the neural tube^{5–7} has helped to change this view. The work of Lee *et al.*³ and others^{8,9} provides evidence for a sequential inductive pathway, whereby the epidermal ectoderm has an initial role in dorsal patterning and neural-crest formation and also induces the roof plate, which then goes on to induce dorsal neurons.

Members of the bone morphogenetic protein (BMP) family of secreted proteins are implicated in mediating signalling by both the epidermal ectoderm and the roof plate, and these tissues express different combinations of BMPs over time⁸. By replacing one copy of the roof-plate-restricted *Gdf7* gene (itself a BMP family member⁹) by the gene for the diphtheria toxin A (DTA) subunit, so that the toxin would be expressed only in roof-plate cells, Lee *et al.*³ genetically ablated the roof plate. This caused the loss of two types of dorsal interneurons, called D1 and D2, and an expansion of more ventral D3 cell types. Lineage analysis of the *Gdf7*-expressing cells shows that the dorsal neurons are not descendants of the roof plate, so they are not missing simply because their progenitors have been eliminated. This has uncovered a non-autonomous role for signals from the roof plate in inducing specific groups of dorsal neurons.

The analysis of the *dreher* mutant adds a new player to this story, moving from outside the cell into the nucleus. Millonig *et al.*⁴ identify *Lmx1a* as the gene mutated or lost in three *dreher* alleles. There are roof-plate defects in the *dreher* mutant, although they are much weaker than in *Gdf7*-DTA embryos. The D2 interneurons are unaffected, D1 interneurons are only partially reduced and the roof-plate defects are restricted to limited regions of the anteroposterior axis. These differences in the severity of dorsal defects in *dreher* might result from residual roof-plate signalling in the affected regions themselves or from the spread of the signal from adjacent unaffected regions. The *dreher* allele studied by Millonig *et al.* contains only a single amino-acid replacement, so it could still have some (albeit reduced) activity. A complete null mutation or other alleles¹⁰ might have a greater effect on dorsal specification. It will be interesting to investigate how the *Lmx1a* and BMP signalling pathways interact.

Developmental biology

Raising the roof

Miguel Manzanares and Robb Krumlauf

In vertebrate embryos, the basic plan of the central nervous system forms as a precise network of neuronal cell types that are generated in ordered patterns in the developing neural tube. Two coordinated but partially independent patterning mechanisms assign positions along the anteroposterior and dorsoventral neural axes^{1,2}. So where a particular cell is generated defines its developmental potential.

Neuronal patterning along the dorsoventral axis appears to involve specific groups of cells or organizing centres that send out opposing signals to tell neighbouring populations what kind of cells to become. Ventral signalling centres such as the floor plate and notochord control motor-neuron patterning. Now, two reports by Lee *et al.*³ and Millonig *et al.*⁴ on pages 734 and 764 of this issue provide new insight into the specification of the most dorsal

neurons. Both confirm that the roof plate, a small group of midline cells, produces signals that induce dorsal types of neurons (Fig. 1). Using elegant genetic ablation and cell-fate-mapping experiments, Lee *et al.*³ show that selective removal of the roof plate in mouse embryos results in loss of the most dorsal interneurons. Millonig *et al.*⁴ have identified the *Lmx1a* gene as the cause of defective roof-plate formation and dorsoventral patterning in the mouse mutant *dreher*. These studies raise a number of questions and concentrate new interest on the roles of the roof plate in dorsal patterning and neurogenesis.

The establishment of dorsoventral polarity of the neural tube depends on a ventral organizing centre. The notochord, a mesodermal rod that runs beneath the neural tube, induces overlying ventral neural tissue to form the floor plate, which can

Daedalus

Down-gazing Dome

London's Millennium Dome is the ultimate triumph of modern styling — all form and no content. Created by the public relations industry, which will say anything but believes nothing, it promises to be an expensive flop. So Daedalus has been thinking of uses for a segment of spherical surface 365 metres in diameter, poised over the surface of the Earth. He reckons it should be ideal for ground-penetrating radar.

This geological technique is widely used for short-range studies. But an accurately shaped downward-pointing aerial 100,000 square metres in area could extend it dramatically. Launching megawatt pulses in the 10-cm to 10-m waveband, it could plumb the geological depths in a way that only seismic shock can do at present.

The Dome will need some modification for the job. The junk that now clutters its interior will have to be removed, and its skin will have to be metallized to reflect and focus the signals. A pit at its central axis, 128 metres deep and 20 metres wide at the bottom, will reach its focal point and allow the focal transmitter-receiver a little sideways movement to steer the beam. The system could then launch accurately collimated radar pulses deep into the Earth, and detect and analyse their echoes.

Conventional seismology detects density or elasticity discontinuities. Dome radar will detect changes in conductance or dielectric character. It should be able to scrutinize the deep slow-flowing regions which generate the geomagnetic field. It might even detect fast transients on the movement, if not the main flow itself, by Doppler effects. By sweeping the beam around, and scanning the wavelength of transmission, the Dome could build up a wonderfully detailed spectrum and map of the Earth's core under London.

Daedalus reckons that the Dome should work as a passive receiver, too. The geomagnetic dynamo, especially if disturbed by seismic shocks, may generate overtones and harmonics which radiate upwards and could be focused and detected by the Dome as a high-gain aerial. And if the New Zealanders can be persuaded to build a duplicate Dome, it might even be possible to detect signals transmitted through the whole Earth, gathering information about its internal structure along a complete diameter. **David Jones**

The Further Inventions of Daedalus (Oxford University Press), 148 past Daedalus columns expanded and illustrated, is now on sale. Special *Nature* offer: m.curtis@nature.com

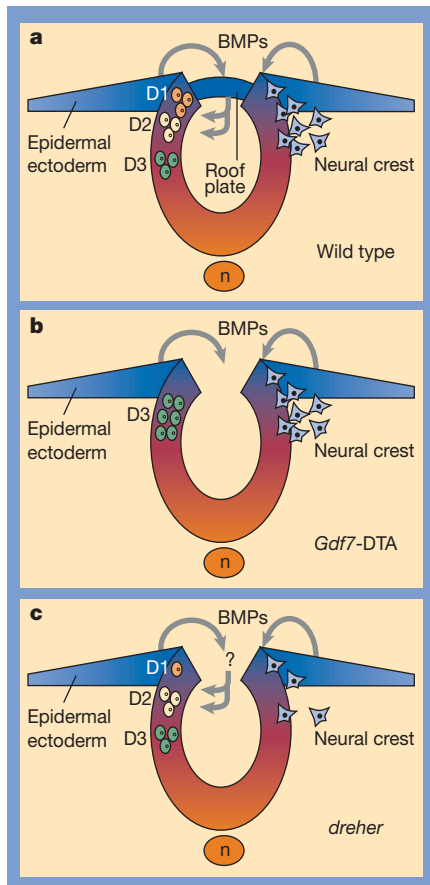


Figure 1 Dorsal signalling in the vertebrate neural tube. **a**, Bone morphogenetic proteins (BMPs) derived from the epidermal ectoderm act on the neural plate to induce first neural crest and then roof plate. The roof plate in turn uses BMPs to signal to the neural ectoderm to induce dorsal types of interneurons (D1 to D3). **b**, Roof-plate ablation in *Gdf7*-DTA embryos results in absence of D1 and D2 neurons, and D3 interneurons expand dorsally³. BMP signalling from the epidermal ectoderm is not sufficient to induce these types of neurons. **c**, In *dreher* mutants there is also no roof plate, but only the specification of D1 interneurons is partially lost⁴. Neural crest derivatives are also affected. It is unknown whether roof-plate signals are lost completely in *dreher* mutants or whether there is a means of compensating for their loss in these mice. **n**, notochord.

How does specification of the roof plate itself relate to that of other dorsally derived structures, such as the neural crest and the rhombic lip (the source of the granule cells of the cerebellum)? The neural crest is normally determined before induction of the roof plate, during a narrow time window⁵. Lee *et al.*³ have shown that many neural-crest derivatives are generated in *Gdf7*-DTA embryos and that epidermal ectoderm maintains its ability to induce dorsal character even when the roof plate is defective. This implies that epidermal ectoderm can induce roof plate and other dorsal

phenotypes (neural crest⁶) through BMP signalling^{7,8,11}, but it does not normally induce differentiated neuronal phenotypes on its own — this job is reserved for the roof plate.

These studies imply that dorsal specification in the neural tube is achieved in two independent and sequential events (Fig. 1). BMP-mediated signals from the epidermal ectoderm to the neural plate induce a general dorsal character in the neural plate that would include first neural crest and then roof plate. In the second phase, the generation of different dorsal cell types would be under the specific influence of the roof plate and later the rhombic lip, independent of the epidermal ectoderm. BMPs act in both of these processes, but the question remains whether they are different mechanisms, or whether the roof plate acts as a relay centre to pass on the epidermal signal. Evidence that functions are divided between different genes, and are not part of a general combinatorial cascade, arises from the observation that only a specific subset of D1 interneurons are affected in the *Gdf7* mutant mice⁹. Finding out which individual BMPs, receptors and associated proteins do the job in each case will be of great interest.

Finally, what other dorsal and ventral signalling centres interact with the floor plate and roof plate? The fact that not all dorsal interneurons are affected by roof-plate ablation (the D3 interneurons are still present in *Gdf7*-DTA mice³) indicates that there must be additional inputs into dorsal patterning. We will have to wait for a compound *Gdf7*-DTA (or *dreher* null)/*Shh* mutant to see what happens in the absence of both dorsal and ventral organizers, if there is anything left at all.

Miguel Manzanares is at the Instituto Cajal, CSIC, Av. Doctor Arce 37, 28002 Madrid, Spain. e-mail: manzanares@cajal.csic.es

Robb Krumlauf is in the Division of Developmental Neurobiology, National Institute for Medical Research, Medical Research Council, The Ridgeway, Mill Hill, London NW7 1AA, UK.

e-mail: rkrumlauf@nimr.mrc.ac.uk

1. Lumsden, A. & Krumlauf, R. *Science* **274**, 1109–1115 (1996).
2. Tanabe, Y. & Jessell, T. M. *Science* **274**, 1115–1123 (1996).
3. Lee, K. J., Dietrich, P. & Jessell, T. M. *Nature* **403**, 734–740 (2000).
4. Millonig, J. H., Millen, K. J. & Hatten, M. E. *Nature* **403**, 764–769 (2000).
5. Selleck, M. A. J. & Bronner-Fraser, M. *Development* **121**, 525–538 (1995).
6. Dickinson, M. E., Selleck, M. A. J., McMahon, A. P. & Bronner-Fraser, M. *Development* **121**, 2099–2106 (1995).
7. Liem, K. F. Jr, Tremml, G., Roelink, H. & Jessell, T. M. *Cell* **82**, 969–979 (1995).
8. Liem, K. F. Jr, Tremml, G. & Jessell, T. M. *Cell* **91**, 127–138 (1997).
9. Lee, K. J., Mendelsohn, M. & Jessell, T. M. *Genes Dev.* **12**, 3394–3407 (1998).
10. Bergstrom, D. E., Gagnon, L. H. & Eicher, E. M. *Genomics* **59**, 291–299 (1999).
11. Alder, J., Lee, K. J., Jessell, T. M. & Hatten, M. E. *Nature Neurosci.* **2**, 535–540 (1999).

Obituary

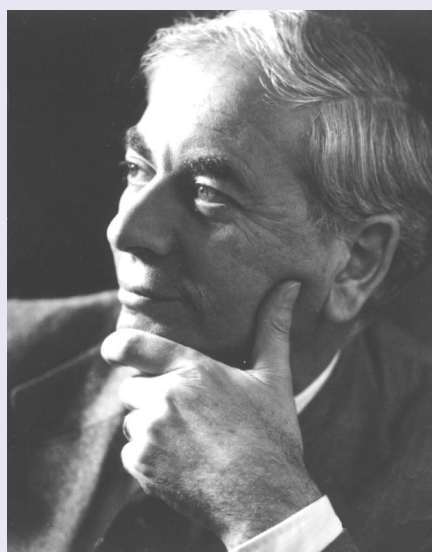
Dennis Sciama (1926–99)

In the years after the Second World War, general relativity theory and cosmology were transformed. Previously seen as only of philosophical interest, they have become central strands in physics, and the new subject of relativistic astrophysics has come into being. Dennis Sciama, who died on 18 December last year, was one of the far-sighted physicists involved in this transition.

Sciama was a student of Paul Dirac, and like him became fascinated with Mach's principle — loosely, the idea that the nature of local physical laws is affected by the state of the whole Universe. Sciama developed a theory of gravity that incorporated this principle, his PhD thesis on the subject being widely quoted. Benefiting from many discussions with Hermann Bondi, Thomas Gold and Fred Hoyle, he developed the broad theme of how local physics relates to the Universe.

Sciama became a passionate believer in the Bondi–Gold–Hoyle steady-state theory (in which the Universe always expands at a constant rate), helping to develop observational tests of that theory. As he emphasized, the theory had the virtue of being disprovable. When it became untenable because of the radio-source and quasar number counts (there are more such sources at large distances than that theory allows), he reluctantly abandoned it. Sciama then became one of the pioneers investigating astrophysical interactions in the evolving and expanding Universe — in particular in studying the way radiation and matter have interacted and how the burgeoning fields of radio and X-ray astronomy could provide information on the thermal history of the Universe.

Here, the transforming event was the discovery in 1965 of the 3 K cosmic blackbody radiation — the relic radiation of the hot big bang — by Arno Penzias and Robert Wilson. This was then tied in to the theory of synthesis of light elements in the early Universe from protons and neutrons, and the theory of growth of large-scale structures (galaxies and clusters of galaxies) in the era after decoupling of matter and the background radiation. The resulting inhomogeneities caused anisotropies in the cosmic background radiation intensity measured today. With Martin Rees, Sciama helped develop the theory of these anisotropies. His considerable understanding of the interlocking set of astrophysical interactions in the early Universe was summarized in *Modern Cosmology*



A force in astrophysics and cosmology

(Cambridge Univ. Press, 1971).

Sciama's career moves took in the universities of Cambridge, Oxford and Texas (Austin), and then the International School of Advanced Studies (SISSA) in Trieste, his interests becoming increasingly broad. They encompassed the structure of radio sources and quasars, X-ray astronomy, the physics of the interstellar and intergalactic medium, astroparticle physics, and the nature of the mysterious dark matter that pervades the Universe, as well as the thermodynamics of black holes and the nature of the vacuum in quantum theory. He also worked on gravitational theory. In particular, he investigated the inclusion of torsion in the theory (allowing for a kind of intrinsic spin in the geometry) and developed an intriguing integral formulation of the Einstein field equations.

More recently, Sciama worked hard on a scheme in which the dark matter in the Universe is taken to consist mainly of massive neutrinos which decay with a long half-life. These neutrinos produce photons, which could then be responsible for a range of observable astrophysical phenomena such as ionization of intergalactic gas. He formulated this theory so that it too was eventually falsifiable; it was a sadness for him when observations failed to support it.

The real impact of Sciama's career lay, however, not in his own contributions, impressive as they are, but in his creation

of an enormously influential school of students. Sciama devoted his life to finding high-quality, dedicated students and helping them in their careers. He supervised over 70 PhD students, among them Stephen Hawking, Brandon Carter (formulator of the Anthropic Principle in cosmology), Sir Martin Rees, Philip Candelas, John Barrow and David Deutsch (originator of quantum computing). Students of students number well over 180, including such luminaries as Roger Blandford, Craig Hogan, Nick Kaiser and Peter Coles. A 'family tree' of this impressive group of relativists and astrophysicists is given in the Sciama Festschrift, *The Renaissance of General Relativity and Cosmology*, edited by myself and colleagues (Cambridge Univ. Press, 1993).

Sciama inculcated in his students the importance of physical and astrophysical understanding, the significance of rigorous mathematical analysis whenever this is possible, and the power of combining the two in a way leading to testable predictions. He encouraged them to be both adventurous and rigorous, and to make their thinking relevant to the physical problem at hand.

And as well as having his own students, he was a proponent of relativistic astrophysics and cosmology wherever he went. Indeed it was Sciama who interested Roger Penrose in gravitation theory when they were together at Cambridge, thereby opening the way to the systematic study of the causal properties of space-times and the famous Penrose–Hawking singularity theorems. These showed that (according to classical general relativity theory) there is indeed a beginning to the Universe, as well as an end to space-time when a massive object collapses to a black hole.

Sciama was a warm person with a love for the civilized things in life. He enjoyed good company, music, and opera, in the later part of his life often retreating to his flat in Venice where, with his wife Lydia, he enjoyed the Venetian lifestyle. When at work he always displayed an intellectual determination to get to grips with the issues concerned, bringing a formidable knowledge of physics and mathematics to bear in this endeavour. His forte was working on these problems with his students and colleagues. He was like a father to many of them, and will be missed both personally and as an inspirational figure in the field of study he loved so much.

George F. R. Ellis

George F. R. Ellis is in the Department of Mathematics, University of Cape Town, Rondebosch 7700, Cape Town, South Africa. e-mail: ellis@maths.uct.ac.za

Conversion of diploidy to haploidy

Individuals susceptible to multigene disorders may now be spotted more easily.

The problem with humans, at least from the perspective of genetic diagnostics, is that they have two copies of each of their chromosomes (diploidy). Mutations in one copy of a chromosome pair can therefore be obscured by the normal sequence present on the other copy of the chromosome. Here we describe a way to overcome this problem and expose the masked mutation by converting the human chromosome complement to a haploid state through fusion to a novel recipient cell line. This approach can significantly increase the sensitivity of genetic tests.

The conversion approach employs fusion between human and rodent cells to create hybrids that contain only a subset of the human chromosomes. Such fusions have proved useful for a variety of purposes, including genetic diagnosis^{1–8}. Investigation of multigenic syndromes, such as those associated with many neoplastic, neurological and cardiovascular diseases, requires the analysis of multiple genes located on different chromosomes. We have devised a robust system involving a new recipient cell line for generating stable hybrids that contain any desired chromosome after a single fusion with a universal recipient.

The features of this recipient cell line (E2) and how it is used to produce hybrids are outlined in Fig. 1. Lymphocytes from 14 individuals were used to produce 476 hybrids to validate the procedure. On average, 28% of the hybrids were found to contain a single copy of each of five different chromosomes tested. Fluorescence *in situ* hybridization (FISH) on metaphase spreads from the hybrids confirmed that whole chromosomes were retained, and indicated the presence of 11 ± 3 human chromosomes in each hybrid cell. The human chromosome complements of the hybrids were remarkably stable, with nearly identical patterns of retention after growth for 90

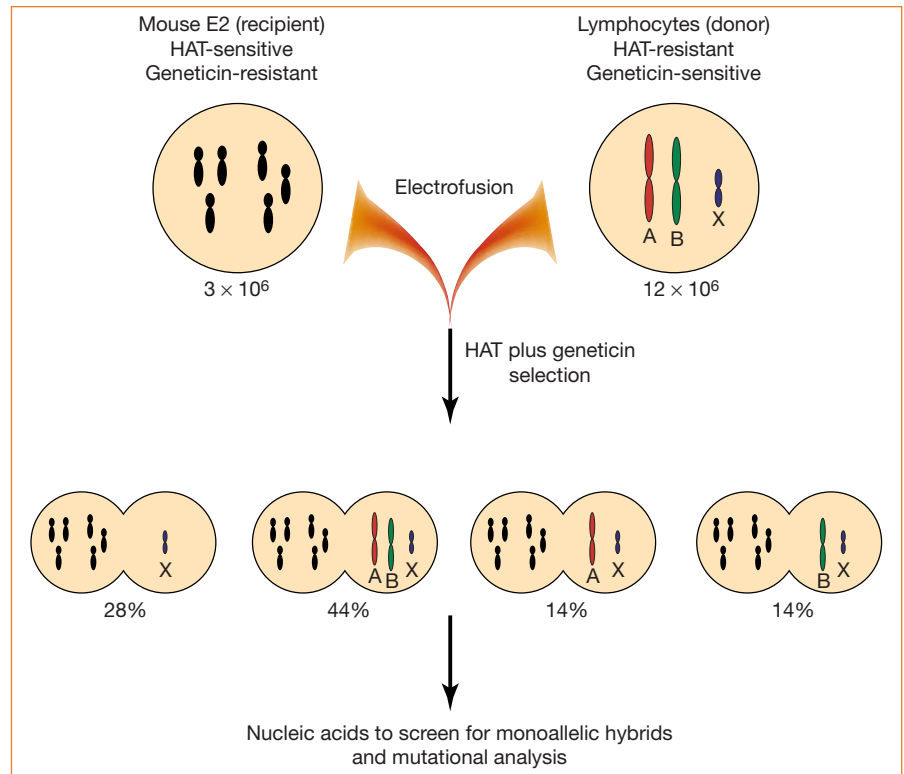


Figure 1 Outline of the conversion approach. To maximize the stability of transferred chromosomes¹³, mouse recipient cells were generated from mouse embryonic fibroblasts derived from MSH2-deficient mice¹⁴. Immortalization was achieved through transformation with adenovirus *E1A* and *ras* oncogenes, and a mutation in the *HPRT* gene made the cells sensitive to HAT (hypoxanthine, aminopterin, thymidine). One of the clones generated, E2, was selected for its growth and fusion characteristics, as well as for its stable diploid chromosomal content. To produce hybrids with E2 cells, human lymphocytes are electrofused with them and selected in HAT plus geneticin, which kills unfused E2 cells and unfused lymphocytes, respectively. Colonies appearing after two weeks of growth are expanded for an additional two weeks, when nucleic acids are prepared for analysis. DNA polymorphisms are used to identify hybrids containing the desired chromosomes. The average proportion of clones that retained neither, both or one homologue of the five chromosomes analysed is indicated. Chromosomes marked A and B represent the two homologues of a representative human chromosome. RNA from selected hybrids is then used to determine whether both alleles of a tested gene are transcribed at equivalent levels and whether the sequences of the encoded proteins are normal. Detailed methods for generation, maintenance and analysis of hybrids are available from the authors.

generations after initial genotyping. Moreover, appropriate hybrids expressed every human gene tested, including the eight known susceptibility genes for colorectal cancer on chromosomes 2, 3, 5, 7 and 16.

To illustrate the value of conversion, we investigated a cohort of 22 patients who had hereditary non-polyposis colorectal cancer (HNPCC) and evidence of mismatch-repair (MMR) deficiency in their tumours⁹. Using conventional techniques¹⁰, we were unable to identify MMR gene mutations in 10 of these 22 patients. Conversion allowed us to identify disease-causing alterations in all 22 patients (Table 1). Three of the cases were due to large deletions of the *hMSH2* gene. Only the normal sequences inherited from the unaffected parent were amplified from the diploid cells of these patients, explaining the failure of conventional approaches based on the polymerase chain reaction to reveal the defects.

In three other cases, no *hMLH1* transcript was generated from one allele, although the sequences of all exons and intron–exon borders from this allele were

Table 1 Summary of MMR gene mutations identified by conversion

Kindred	Gene	Functional abnormality	Genomic DNA alteration
1	MSH2	No detectable transcript	Deletion of exons 1–6
2	MSH2	R680Ter	CGA to TGA
3	MSH2	No detectable transcript	Deletion of exons 1–7
4	MLH1	No detectable transcript	NI
5	MSH2	A636P*	GCA to CCA
6	MLH1	No detectable transcript	NI
7	MSH2	R680Ter	CGA to TGA
8	MSH2	No detectable transcript	Deletion of exons 1–7
9	MLH1	No detectable transcript	NI
10	MSH2	24-bp insertion between codons 215 and 216	TAG to GAG, splice acceptor of exon 4

NI, genomic alteration not identified, despite complete sequencing of all exons and intron–exon borders of the abnormal allele.

*Found in other HNPCC kindreds, but not in 400 normal alleles (ref. 15 and our unpublished data); mutant product functionally defective¹⁵.

normal. Presumably, mutations deep within introns or in the promoter of the *hMLH1* gene were responsible. In four other cases, we identified point mutations in the *hMSH2* gene. These mutations were not detected in analogous templates from diploid cells because the signals from the mutant allele were weaker than those from the normal allele. Such asymmetry is a common problem for both manual and automated sequencing methods, and is eliminated through conversion because only one sequence can be present at each nucleotide position.

This haploid conversion thus allowed us to increase significantly the sensitivity of genetic testing for HNPCC. By combining several techniques (for example, western blotting, protein-truncation tests, sequencing of genomic DNA as well as of complementary DNA, Southern blotting and FISH), the sensitivity of conventional genetic testing can be increased¹¹, but such technical combinations are impractical for analysing multigenic diseases. It is interesting that one of the patients in our cohort was a member of Warthin's family G (ref. 12): description of this family initiated research into colorectal cancer genetics in 1913, but the mutation responsible had eluded intensive efforts to identify it by conventional techniques.

Diploid-to-haploid conversion is not a substitute for the many powerful methods that are currently available to search for mutations. Rather, conversion can be used to maximize the sensitivity of such techniques. Converted nucleic acids are the preferred substrates for these methods because of the much higher signal-to-noise ratio that is attainable and because the wild-type allele cannot confound detection of the mutant allele. As mutational assays improve through the incorporation of microarrays and other automatable features, the value of conversion-generated nucleic acids should increase correspondingly, markedly enhancing the effectiveness of genetic tests for hereditary diseases.

Hai Yan*, Nickolas Papadopoulos†, Giancarlo Marra‡, Claudia Perrera‡, Josef Jiricny‡, C. Richard Boland§, Henry T. Lynch||, Robert B. Chadwick¶, Albert de la Chapelle¶, Karin Berg*, James R. Eshleman*, Weishi Yuan*, Sanford Markowitz#, Steven J. Laken*, Christoph Lengauer*, Kenneth W. Kinzler*, Bert Vogelstein*

*Howard Hughes Medical Institute, The Oncology Center, Department of Mathematical Sciences, and the Department of Pathology, The Johns Hopkins University, Baltimore, Maryland 21231, USA
e-mail: vogelbe@welchlink.welch.jhu.edu

†Institute of Cancer Genetics, Columbia University, New York, New York 10032, USA

‡Institute of Medical Radiobiology, August Forel-Strasse 7, CH-8008 Zurich, Switzerland

§University of California, San Diego, La Jolla, California 92093, USA

||Department of Preventive Medicine and Public Health, Creighton University School of Medicine, Omaha, Nebraska 68178, USA

¶Human Cancer Genetics Program, Comprehensive Cancer Center, Ohio State University, Columbus, Ohio 43210, USA

#Howard Hughes Medical Institute, and Department of Medicine and Ireland Cancer Center, Case Western Reserve University and University Hospitals of Cleveland, Cleveland, Ohio 44106, USA

1. Paterson, D., Jones, C., Scoggin, C., Miller, Y. E. & Graw, S. *Ann. NY Acad. Sci.* **396**, 69–81 (1982).

Epidemiology

Reactivation of *Borrelia* infection in birds

Birds often carry ticks infected with *Borrelia burgdorferi sensu lato*¹ — the spirochaete that causes Lyme disease — but their role as possible hosts and amplifiers for this illness has long been discounted. We find, however, that migratory birds are able to carry Lyme disease as a latent infection for several months, and that this infection can be reactivated and passed on to ticks as a result of migratory restlessness. Our results indicate that migratory birds may be efficient long-distance carriers of this pathogen.

Ixodid ticks are vectors of *B. burgdorferi* s.l. spirochaetes that cause Lyme disease. At least three genospecies within the *B. burgdorferi* s.l. group are pathogenic to humans: *B. burgdorferi sensu stricto*, *B. afzelii* and *B. garinii*. We tested whether birds could carry *B. garinii* and reactivate it under stressful conditions by simulating their migration.

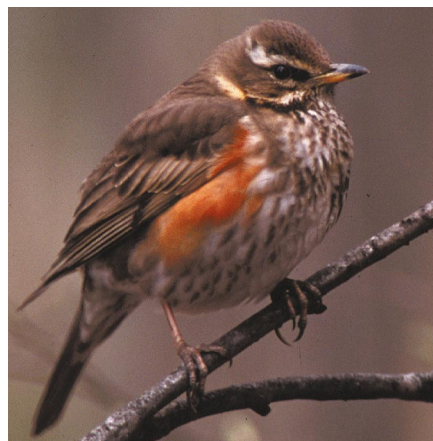


Figure 1 The redwing thrush (*Turdus iliacus*), a common migratory bird that is often infested with ticks, can carry Lyme disease as a latent infection for several months. These infections can be subsequently reactivated by migratory restlessness and spread great distances during the bird's migratory travels.

- Harris, H. J. *Cell Sci.* **79**, 83–94 (1985).
- Schultz, R. A., Saxon, P. J., Glover, T. W. & Friedberg, E. C. *Proc. Natl Acad. Sci. USA* **84**, 4176–4179 (1987).
- Grodin, J. et al. *Cell* **66**, 589–600 (1991).
- Anderson, M. J. & Stanbridge, E. J. *FASEB J.* **7**, 826–833 (1993).
- Imreh, S. et al. *Genes Chromosom. Cancer* **11**, 237–245 (1994).
- Papadopoulos, N., Leach, F. S., Kinzler, K. W. & Vogelstein, B. *Nature Genet.* **11**, 99–102 (1995).
- McDaniel, L. D., Legerski, R., Lehmann, A. R., Friedberg, E. C. & Schultz, R. A. *Hum. Mutat.* **10**, 317–321 (1997).
- Lynch, H. T. & de la Chapelle, A. *J. Med. Genet.* **36**, 801–818 (1999).
- Liu, B. et al. *Nature Med.* **2**, 169–174 (1996).
- Wijnen, J. et al. *Nature Genet.* **20**, 326–328 (1998).
- Warthin, A. S. *Arch. Int. Med.* **12**, 546–555 (1913).
- Lengauer, C., Kinzler, K. W. & Vogelstein, B. *Nature* **386**, 623–627 (1997).
- Reitmair, A. H. et al. *Nature Genet.* **11**, 64–70 (1995).
- Yuan, Z. Q. et al. *J. Med. Genet.* **36**, 790–793 (1999).

Stress, especially strenuous exercise, suppresses immune systems^{2,3}. Hormonal regulation of the immune system, particularly by increased circulating glucocorticoids, may also adversely affect immunocompetence⁴ and activate latent infections⁵. As these stress hormones are increased in birds during migration⁶, we investigated whether the stress of migration could impair a bird's defence against *B. garinii* and so increase the degree of spirochaetemia. We chose redwing thrushes for our study because these are an abundant migratory species frequently infested by the ticks that transmit Lyme disease *Borrelia*¹.

Nineteen juvenile redwing thrushes (Fig. 1) were divided into two groups and housed in separate rooms. One room (the control room) had a constant 12-hour photoperiod. The other room (the migration room) had its daily photoperiod gradually decreased from 12 to 3.5 hours over the first two months of the three-month experiment to simulate the decreasing daylight of autumn. After day 50, this shorter photoperiod triggered increased nocturnal activity in the thrushes (Mann–Whitney *U*-test, *P* = 0.001; Fig. 2a), a behaviour called migratory restlessness. Inducing this behaviour and its accompanying physiological condition is a widely accepted method for investigating features of the migration process in the laboratory⁷, which assumes that preventing birds from migrating does not confound the results.

Up until day 49, there was no difference in body mass between the two groups of birds. Body mass increased steadily, as is typical of premigratory fat deposition⁷. But during days 64–93, body mass remained high in the control room, but decreased in the migration room (Mann–Whitney *U*-test, *P* = 0.008; Fig. 2b), probably because of the increased nocturnal activity.

Eight of ten birds in the migration room, and six of nine in the control room, were infected on day zero with a subcutaneous injection of 10⁴ spirochaetes of *B. garinii*. From days 71 to 87, viable

spirochaetes were detected in cultures from blood samples and subcutaneous aspirates from five of the infected birds in the migration room. No spirochaetes were found in samples taken before that time, nor in those from birds in the control room. There was a significantly higher proportion of culture-positive birds in the migration room (five out of eight) than in the control room (zero out of six) (Pearson $\chi^2=5.833$, 1 d.f., $P=0.031$; Fig. 2c). Sequencing a variable part of the *ospA* gene from all cultivated isolates confirmed that the sequences were identical to that of the infecting strain.

As Lyme disease *Borrelia* grows optimally *in vitro* at 34–37 °C (ref. 8), the high mean body temperature of passerine birds (40.6 °C; ref. 9) was considered to rule out their likelihood as amplification hosts. The body temperature of birds is not uniform, however, either spatially or temporally, and some parts (skin and air sacs, for example) may be at lower temperatures than the

internal organs⁹. Of the three different *Borrelia* species causing Lyme disease, *B. garinii* seems to have the highest temperature for optimal growth¹⁰. In addition, *B. garinii* and *B. valaisiana*, which does not cause disease in humans, are the species most often found in birds, suggesting that these are well adapted to higher body temperatures^{11,12}.

Our results show that migratory restlessness in redwing thrushes can reactivate a latent *Borrelia* infection. As thrushes and other birds often travel great distances during migration, this reveals a new mechanism for facilitating the long-distance spread of Lyme disease. Ticks anywhere along a migration route can feed on migrants with reactivated infections and become infected themselves, in turn passing the disease on to other organisms.

Åsa Gylfe*, Sven Bergström*, Jan Lundström†, Björn Olsen‡§

*Department of Microbiology, Umeå University, SE-901 87 Umeå, Sweden

†Department of Population Biology, Evolutionary Biology Center, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

‡Department of Infectious Diseases, Umeå University, SE-901 87 Umeå, Sweden

§Department of Infectious Diseases, Kalmar County Hospital, SE-391 85 Kalmar, Sweden

e-mail: bjornoln@ltkalmars.se

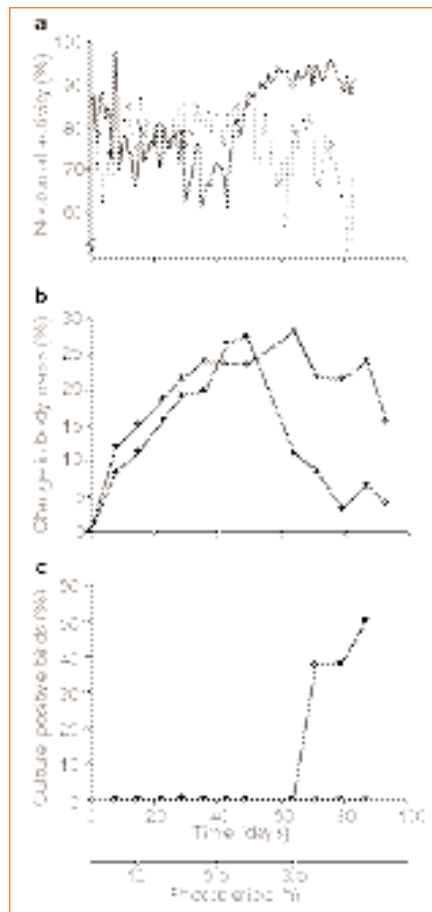


Figure 2 *Borrelia garinii* infection in redwing thrushes. **a**, Nocturnal activity; **b**, change in body mass; and **c**, reactivation of infection in these birds. Photoperiod in the migration room is indicated (bottom). Solid line/filled points, migration room; dashed line/open points, control room. The juvenile redwing thrushes were caught in an area where ticks and Lyme disease are rare. Each bird was kept in an individual cage (57 × 37 × 50 cm) equipped with an infrared sensor that registered all movement such as wing whirring, fluttering and hopping. *Borrelia* spirochaetes were cultured as described⁸.

- Olsen, B., Jaenson, T. G. T. & Bergström, S. *Appl. Environ. Microbiol.* **61**, 3082–3087 (1995).
- Räberg, L., Grahn, M., Hasselquist, D. & Svensson, E. *Proc. R. Soc. Lond. B* **265**, 1637–1641 (1998).
- Apanius, V. in *Stress and Behaviour* (eds Møller, A. P., Milinski, M. & Slater, P. J. B.) 133–153 (Academic, San Diego, 1998).
- Besedovsky, H. O. & del Rey, A. *Endocr. Rev.* **17**, 64–102 (1996).
- Mackowiak, P. A. *Rev. Infect. Dis.* **6**, 649–668 (1984).
- Holberton, R. L., Parrish, J. D. & Wingfield, J. C. *Auk* **113**, 558–564 (1996).
- Alerstam, T. *Bird Migration* (Cambridge Univ. Press, 1990).
- Barbour, A. G. *Yale J. Biol. Med.* **57**, 521–525 (1984).
- Welty, J. C. & Baptista, L. in *The Life of Birds* 119–141 (Saunders College Publishing, New York, 1988).
- Hubálek, Z., Halouzka, J. & Heroldová, M. *J. Med. Microbiol.* **47**, 929–932 (1998).
- Humair, P. F. *et al. Zentralbl. Bakteri.* **287**, 521–538 (1998).
- Kurtenbach, K. *et al. Appl. Environ. Microbiol.* **64**, 1169–1174 (1998).

Marine ecology

Do mussels take wooden steps to deep-sea vents?

Symbiont-containing mussels (Mytilidae) are found at hydrothermal vents and cold seeps on the ocean floor, but it is not known whether these taxa represent an ancient lineage endemic to these surroundings or are more recent invaders. Here we show that several small and poorly known mussels, commonly found on sunken wood and whale bones in the deep sea, are closely related to vent and seep taxa, and that this entire group is divergent from other Mytilidae. Our results indicate that vents and seeps were recently invaded by

modern mytilid taxa and suggest that decomposing wood and bone may have served as ‘steps’ for the introduction of mytilid taxa to vents and seeps.

Symbiont-containing mussels from hydrothermal vents and cold seeps are currently placed in the recently proposed mytilid subfamily Bathymodiolinae¹. This subfamily contains 11 species in two genera, *Bathymodiolus* and *Tamu*, all of which obtain nutrients fixed by chemoautotrophic or methanotrophic endosymbionts harboured inside markedly enlarged gills. Most of these mussels are large and all are indigenous to deep-sea hydrothermal vents or cold-water seeps².

Chemoautotrophic endosymbionts have been discovered in the gills of a deep-sea mytilid not typically found in vent and seep environments³. This tiny (average length ~4 mm) mussel, *Idas washingtonia*, was the most abundant species (more than 10,000 individuals per site) in chemoautotrophic invertebrate communities living on four lipid-rich whale skeletons at depths of 1,000–2,000 m on the northeast Pacific slope^{4,5}. Hydrogen sulphide, produced by bacterial decomposition of lipids in the bones, rather than vent effluents⁴, probably supports these communities⁵.

Idas washingtonia also has an unexplained habit of wood association. It belongs to one of several genera of small mussels of unknown phylogenetic affinity commonly found on sunken wood, woody plant materials and whale bones at depths from 150 to >3,500 m on the sea floor⁶.

Surprisingly, our phylogenetic analyses based on comparison of gene sequences encoding ribosomal RNA (18S rRNA) indicate that all our samples of wood and bone mussels form a monophyletic lineage that includes all examined vent and seep species, but excludes members of traditional mytilid subfamilies (Fig. 1). These analyses provide strong support for the proposed subfamily status of Bathymodiolinae but suggest that the wood- and bone-associated genera *Adipicola*, *Idas*, *Myrina* and *Benthomodiolus* should also be included in this subfamily.

The almost identical 18S rRNA sequences within this clade (>98.5% for the entire group, >99.7% among vent and seep taxa) are consistent with recent diversification and recent invasion of vent and seep environments. Although internal branching order is not fully resolved, the basal divergence of *Benthomodiolus lignicola*, a species observed on sunken wood and bone, receives significant bootstrap support⁷ (77–89%) by all inference methods, suggesting that wood and bone association may have preceded vent and seep specialization within this lineage.

The association of *I. washingtonia* with sunken wood can now be re-evaluated. As in the case of whale bones, sulphide is a

principal product of the bacterial degradation of wood in marine environments⁸. Thus, *Idas*, through its chemosynthetic endosymbionts, may also use sulphide produced by wood decomposition.

Considering their close phylogenetic relationship and similar habitat preference, we predict that similar chemoautotrophic symbioses exist in other wood- and bone-associated mussels. Wood and woody plant materials, like whale bones, sustain other taxa known to harbour chemoautotrophic endosymbionts, including thyasirid⁹ and

solemyid clams¹⁰, and vestimentiferan tube-worms¹¹. In fact, vestimentiferans have been reported in only two non-vent or seep settings: a whale fall in the Santa Catalina Basin¹² and woody plant materials (bales of sisal twine near sacks of beans and seeds) in a sunken freighter in the eastern Atlantic¹¹; in both cases, the tubeworms occurred together with *Idas* sp.

Why should decomposing wood and bone support vent-like, chemosynthesis-based invertebrate communities whereas other organic substrates typically do not?

Both wood and bone have a high reduced-carbon content, resist consumption by detritivores, and undergo slow and sustained decay. Thus, whereas opportunistic feeders rapidly consume most types of organic detritus in the deep sea⁹, large wood and bone deposits may require more than ten years to decompose^{4,5,13}. Such sustained decomposition could be necessary to support chemoautotrophy-dependent species, which may require several years to reproduce⁵.

Whale falls have been proposed to act as evolutionary stepping stones for the introduction of chemoautotrophy-dependent invertebrates to vent and seep environments⁵. Sunken wood and woody plant materials may act similarly. These substrates are common in the deep sea¹⁴ and probably have been since the Carboniferous, long before bivalve fossils appeared at vents or seeps¹⁵. The close phylogenetic relationships among wood-, bone-, vent- and seep-associated mussels are indicative of a recent common ancestry for vent and non-vent species, whereas the divergence of the entire group from other mytilids is indicative of a long isolation from common shallow-water taxa. These observations implicate wood and bone as historical vectors for the transport of mytilids to (or from) vents and seeps, and suggest a plausible role for wood- and bone-association as an evolutionary step towards invasion of vent and seep habitats.

Daniel L. Distel*, Amy R. Baco†, Ellie Chuang‡, Wendy Morrill‡, Colleen Cavanaugh‡, Craig R. Smith‡

*Department of Biochemistry, Microbiology, and Molecular Biology, University of Maine, Orono, Maine 04469-5735, USA
e-mail: distel@maine.edu

†Department of Oceanography, University of Hawaii, 1000 Pope Road, Honolulu, Hawaii 96822, USA

‡Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138, USA

1. Kenk, V. C. & Wilson, B. R. *Malacologia* **26**, 253–271 (1985).
2. Gustafson, R. G., Turner, R. D., Lutz, R. A. & Vrijenhoek, R. C. *Malacologia* **40**, 63–112 (1998).
3. Deming, J. W., Reysenbach, A. L., Macko, S. A. & Smith, C. R. *J. Microsc. Res. Tech.* **37**, 162–170 (1997).
4. Bennett, B. A., Smith, C. R., Glaser, B. & Maybaum, H. L. *Mar. Ecol. Prog. Ser.* **108**, 205–223 (1994).
5. Smith, C. R., Kukert, H., Wheatcroft, R. A., Jumars, P. A. & Deming, J. W. *Nature* **341**, 27–28 (1989).
6. Dell, R. K. *Natl Mus. NZ Rec.* **3**, 17–36 (1987).
7. Hillis, D. M. & Bull, J. J. *Syst. Biol.* **42**, 182–192 (1993).
8. Leschine, S. B. *Annu. Rev. Microbiol.* **49**, 399–426 (1995).
9. Grassle, J. F. & Morse-Porteous, L. S. *Deep-Sea Res.* **34**, 1911–1950 (1987).
10. Reid, R. G. B. *Can. J. Zool.* **58**, 386–393 (1980).
11. Dando, P. R., Southward, A. J. & Southward, E. C. *Nature* **356**, 667 (1992).
12. Feldman, R. A., Shank, T. M., Black, M. B., Baco, A. R. & Smith, C. R. *Biol. Bull.* **194**, 116–119 (1998).
13. Dean, H. K. *Malacologia* **35**, 21–41 (1993).
14. Wolff, T. *Sarsia* **1**, 117–136 (1979).
15. Campbell, K. A. & Bottjer, D. J. *Geology* **23**, 321–324 (1995).
16. Swofford, D. L. *PAUP* 4.0 (Phylogenetic Analysis Using Parsimony)* (Sinauer, Sunderland, MA, 1997).
17. Distel, D. L. *Mol. Phylogenet. Evol.* (in the press).

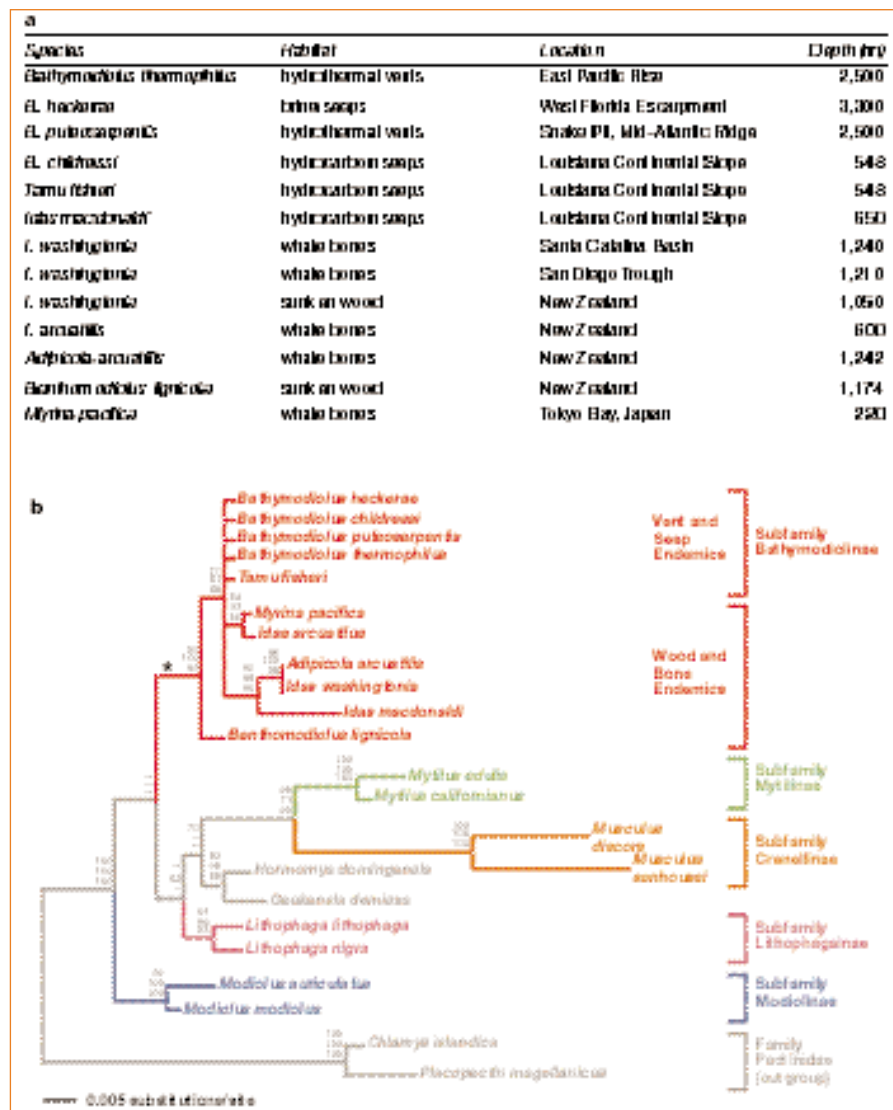


Figure 1 Phylogenetic relationships among vent and non-vent mytilids. **a**, Taxa for which new DNA sequences for 18S rRNA were examined. **b**, Vent-, seep-, wood- and bone-associated mussels form a clade (red), distinct from the traditional subfamilies Mytilinae (green), Crenellinae (brown), Lithophaginae (purple) and Modiolinae (blue). Phylogenetic analyses used maximum likelihood (shown), maximum parsimony and evolutionary distance algorithms¹⁶. Likelihood substitution model: HKY-85 + Γ , with $\alpha = 0.865$ and base frequencies and ti:tv (1.41) estimated from the data. Heuristic searches were done with random sequence addition (100 replicates) and branch swapping by tree-bisection-reconnection. Bootstrap proportions (percentage of 1,000 replicates) are presented (top to bottom: likelihood, distance, parsimony). Dashes are values of <50%. The single best likelihood tree ($-\ln L = 4,328.3$) has topology identical to the best distance and parsimony trees (319 steps) at all nodes for which bootstrap proportions are shown. Asterisk indicates significant support for monophyly determined under Kishino-Hasegawa criteria (best parsimony trees constrained to non-monophyly for the indicated clade are significantly longer ($P < 0.05$) than best unconstrained trees). The aligned sequence set contains 23 taxa, 1,758 equally weighted sites (139 parsimony informative), aligned manually and with consideration of secondary structural information, confirmed by compensatory substitutions with unalignable regions eliminated and gaps treated as missing data (GenBank accession nos AF221638–AF221648). *Hormomya domingensis* and *Geukensia demissa* form a clade distinct from their traditional subfamilies Mytilinae and Modiolinae¹⁷.

Supernova explosions in the Universe

Adam Burrows

Department of Astronomy, The University of Arizona, Tucson, Arizona 85721, USA

During the lifetime of our Milky Way galaxy, there have been something like 100 million supernova explosions, which have enriched the Galaxy with the oxygen we breathe, the iron in our cars, the calcium in our bones and the silicon in the rocks beneath our feet. These exploding stars also influence the birth of new stars and are the source of the energetic cosmic rays that irradiate us on the Earth. The prodigious amount of energy ($\sim 10^{51}$, or $\sim 2.5 \times 10^{28}$ megatonnes of TNT equivalent) and momentum associated with each supernova may even have helped to shape galaxies as they formed in the early Universe. Supernovae are now being used to measure the geometry of the Universe, and have recently been implicated in the decades-old mystery of the origin of the γ -ray bursts. Together with major conceptual advances in our theoretical understanding of supernovae, these developments have made supernovae the centre of attention in astrophysics.

Supernovae are crucial to the dynamical and morphological development of the Universe. They are also at the nexus of many of the great debates now raging among astronomers. One subtype of supernovae, the so-called type Ias, is now arguably astronomy's most accurate probe of the scale and geometry of the Universe. An unknown fraction of another subtype, the core-collapse supernovae, may be the source of γ -ray bursts. As major sources of the elements of existence, supernovae themselves are primary agents of stellar and galactic evolution. Supernovae and γ -ray bursts share the distinction of being the most powerful explosions in the cosmos, and recent observational and theoretical breakthroughs and a renewed appreciation of the manifold roles of supernovae have inaugurated a new era in their study. Here I attempt to summarize, from one theorist's perspective, these new developments. In particular, I survey the context of supernova theory, then delve into the physics of core-collapse supernova explosions, continue by summarizing the emerging connection between supernovae and γ -ray bursts, and finally highlight the role of type Ia supernovae as 'standard candles' with which to measure the geometry and dynamics of the Universe. A clear and inescapable subtext of this review is the centrality of supernovae and their aftermaths to the fundamental questions of astrophysics and cosmology.

About supernovae

It is by its death that the purpose of a massive star is most clearly revealed. All stars are born, have thermonuclear lives, and die, leaving behind tiny fossils. For most stars, including our Sun, these fossils are or will be carbon/oxygen white dwarfs with radii near that of the Earth, masses near 0.5–1.0 solar masses ($M_{\odot}$), and central densities $\sim 10^7$ times that of tungsten. Low-mass stars die and white dwarfs are born slowly over hundreds to thousands of years through the ejection of the dying star's heavy outer mantle. There is no explosion. The dense white-dwarf residue, if in isolation, then cools into obscurity.

In contrast, a star more massive than $\sim 8M_{\odot}$ does not go with a whimper. Without the quietus of gentle mantle ejection, the white dwarf that such a star creates in its core during its last thermonuclear stages continues to evolve in composition and grow in mass and density until it achieves the so-called "Chandrasekhar" mass near $1.4M_{\odot}$ (refs 1,2). At this mass, an iron or oxygen–neon–magnesium white dwarf, normally supported against gravity by electron degeneracy pressure, becomes sufficiently unstable to collapse. Owing to the Pauli exclusion principle, at the high densities achieved by massive white dwarfs, their electrons are relativistic. Unlike a non-relativistic gas, a relativistic gas has a soft equation of state and is easily compressed by the ineluctable forces of persistent gravity. Within one second, the core of a star that may have lived for ten million years, cooking its hydrogen into progressively heavier

elements, implodes from something the size of our planet to something the size of a city, achieving densities in excess of that of the atomic nucleus and velocities one-fourth the speed of light. At nuclear densities ($\sim 10^{13}$ times that of tungsten), matter is barely compressible and the core bounces³, rebounding into the infalling inner mantle and, like a piston, generating a strong shock front that with effort and a short delay overcomes the confining tamp of imploding mantle mass in order to launch a supernova explosion. The violent explosion disassembles the massive star, litters the interstellar medium with freshly synthesized heavy elements (such as oxygen, carbon, magnesium, silicon, calcium, sulphur, and radioactive ^{56}Ni), blows a many-parsec-sized hole in the surrounding galactic gas, and announces itself with a luminous display that can rival that of its parent galaxy for months. Its fossil is most often a neutron star, twenty kilometres wide, with an average density near that of an atomic nucleus, spinning with a period of milliseconds to seconds, and possessing a surface magnetic field of $\sim 10^{12}$ gauss. With the right combination of period and field, this object is a radio pulsar. Astronomers have discovered more than 1,000 such radio beacons in the galaxy, each of which was born in a supernova explosion⁴. The famous Crab pulsar is one such result, now pulsing in the radio, optical, and X-ray frequencies with a period of ~ 30 milliseconds and surrounded by an X-ray emitting remnant of the supernova explosion, that itself was witnessed by humans in AD 1054 to be as bright as Venus in the night sky. Figure 1 depicts the various stages of core collapse and neutron star formation.

The gravitational collapse of the core of a massive star is not the only context in which a supernova explosion is thought to occur. The smaller carbon–oxygen white dwarfs produced by low-mass stars at their death might be situated in tight binaries with a stellar companion. In a small subset of these systems, matter is stripped from this companion by the white dwarf's gravitational pull and accreted at a rate sufficient to reach the Chandrasekhar mass. As with the core of a dying massive star, collapse ensues. However, since these white dwarfs consist predominantly of carbon and oxygen, not heavier elements such as iron near the peak of the nuclear binding energy curve, compression and heating soon lead, not to continued implosion, but to the thermonuclear incineration of the white dwarf. As much as a solar mass burns to iron-peak and intermediate-mass (such as Ca, Si, S, Ne, Mg) elements and the entire star explosively disassembles, leaving nothing behind but its violently disturbed donor⁵. In some models, the donor is another white dwarf, in which case nothing at all remains. In either case, the ejecta are rich in heavy elements, in particular ^{56}Ni , which by its radioactive decay to ^{56}Co (in a mean time of 8.8 days) and then to stable ^{56}Fe (in a mean time of 111.3 days) powers its optical light curve (luminosity versus time) for the months during which it can be seen across the Universe^{6,7}. Astronomers believe these to be the so-called

type Ia supernovae. If it were not for radioactive heating, adiabatic expansion of the debris would cool it to near invisibility in less than an hour. Type Ia supernovae are about ten times less prevalent than core-collapse supernovae, but yield about ten times as much iron, are often more than ten times brighter at peak light, and are spectacular sources of nuclear γ -ray lines and continuum⁸. It is with these bright supernovae that observers are now obtaining the best and, perhaps, the most provocative information about the geometry of the Universe.

Astronomers use observational, not theoretical, criteria to type supernovae. A type I supernova (such as a type Ia) is one with no hydrogen in its spectrum, while the spectrum of a type II supernova has prominent hydrogen lines. The epochal supernova in the Large Magellanic Cloud (LMC), SN1987A, was a core-collapse supernova, because it exploded as a $\sim 15\text{--}20M_{\odot}$ blue supergiant with a radius of $\sim 4 \times 10^7$ km (ref. 9) and not as the canonical red supergiant with a

radius of $\sim 10^9$ km; however, it was dimmer than a typical type II and early relied on ^{56}Ni to power its muted optical light curve. Yet there is no reason to suspect that the explosion itself was not of the common core-collapse variety. The light curve and spectrum of a supernova reflect more its progenitor's radius, chemical makeup, and expansion velocities than the mechanism by which it exploded. To the theorist, the achievement of the critical Chandrasekhar mass unites the types; the supernova mechanism is either by implosion to nuclear densities and subsequent hydrodynamic ejection, or by thermonuclear runaway and explosive incineration.

There is approximately one supernova explosion in the Universe every second. In our galaxy, there is one supernova every $\sim 30\text{--}50$ years and one type Ia supernova every ~ 300 years. Supernova hunters, peering deeply with only modest-aperture telescopes, can now capture a dozen or so extragalactic supernovae per night, mostly the bright type Ias. Approximately 200 supernova remnant shells are known in the Milky Way and these are radio, optical, and X-ray echoes of only the most recent galactic supernova explosions. Within the last millennium, humans have witnessed and recorded six supernovae in our galaxy (Table 1).

Supernovae from massive stars

A star's first thermonuclear stage is the fusion of hydrogen into helium in its hot core. With the exhaustion of core hydrogen, most stars then proceed to shell hydrogen burning, and then to core helium burning. The ashes of the latter are predominantly carbon and oxygen and low-mass stars do not proceed beyond this stage. However, stars with masses from $\sim 8M_{\odot}$ to $\sim 60\text{--}100M_{\odot}$ (the upper limit depending upon the heavy-element fraction at birth) proceed to carbon burning, with mostly oxygen, neon, and magnesium as ashes^{1,2}. For stars more massive than $\sim 9\text{--}10M_{\odot}$, the ashes of carbon burning achieve sufficient temperatures to ignite and they burn predominantly to silicon, sulphur, calcium, and argon. Finally, these products ignite to produce iron and its congener isotopes near the peak of the nuclear binding energy curve. Fusion is exothermic only for the assembly of lighter elements into elements up to the iron group, not beyond. Hence, at the end of a massive star's thermonuclear life, it has an 'onion-skin' structure in which an iron or oxygen–neon–magnesium core is nested within shells comprised of elements of progressively lower atomic weight at progressively lower densities and temperatures. The outer zone consists of unburned hydrogen and 'primordial' helium. A typical nesting is $\text{Fe} \rightarrow \text{Si} \rightarrow \text{O} \rightarrow \text{He} \rightarrow \text{H}$. The oxygen in the 'oxygen' zone is the major source of oxygen in the Universe, for little oxygen survives in the ejecta of the rarer type Ia supernovae. These shells are not pure,

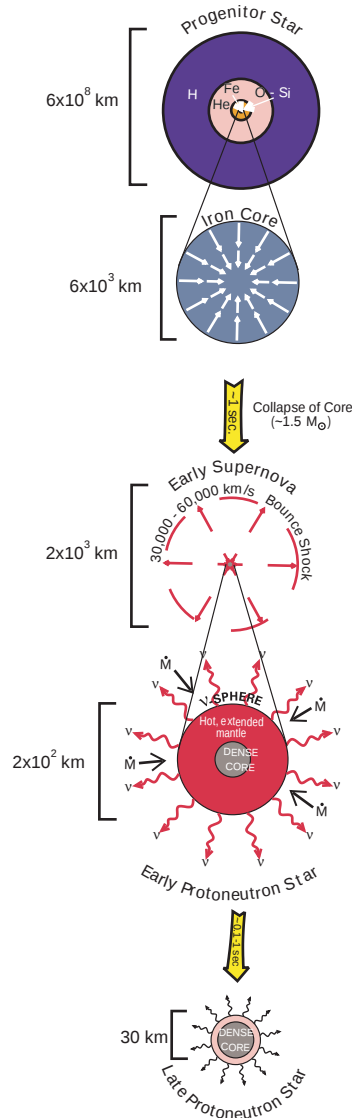


Figure 1 The sequence of events in the collapse of a stellar core to a nascent neutron star. It begins with a massive star with an 'onion-skin' structure, goes through white-dwarf core implosion, to core bounce and shock-wave formation, to the protoneutron-star stage before explosion, and finally to the cooling and isolated-neutron-star stage after explosion. This figure is not to scale. The wavy arrows depict escaping neutrinos and the straight arrows depict mass motion.

Table 1 Supernovae that have exploded in our Galaxy and the Large Magellanic Cloud within the last millennium

Supernova	Year (AD)	Distance (kpc)	Peak visual magnitude
SN1006	1006	2.0	−9.0
Crab	1054	2.2	−4.0
SN1181	1181	8.0	?
RX J0852-4642	~1300	~0.2	?
Tycho	1572	7.0	−4.0
Kepler	1604	10.0	−3.0
Cas A	~1680	3.4	~−6.0?
SN1987A	1987	50 ± 5	3.0

These 'historical' supernovae are only a fraction of the total, because the majority were shrouded from view by the dust that pervades the Milky Way. Thus, it is estimated that this historical cohort represents only about 20% of the galactic supernovae that exploded since AD1000. Included are SN1987A, which exploded not in the Milky Way but in the Large Magellanic Cloud (one of its nearby satellite galaxies), RX J0852-4642 (ref. 77, ref. 11), a supernova remnant whose recent ($\sim \text{AD}1300$) and very nearby birth went unrecorded, perhaps because it resides in the Southern Hemisphere (but in fact for reasons that are as yet unknown), and Cas A, a supernova remnant that was born in historical times, but whose fiery birth was accompanied by a muted visual display that may have been recorded only in the ambiguous notes of the seventeenth-century astronomer John Flamsteed (ref. 78). The distances and peak visual magnitudes quoted are guesses at best, except for SN1987A. Astronomical magnitudes are logarithmic and are given by the formula $M_V = -2.5\log_{10}(\text{brightness}) + \text{constant}$. Hence, every factor of ten increase in brightness represents a decrease in magnitude by 2.5. For comparison, the Moon is near -12 magnitudes, Venus at peak is -4.4 magnitudes, and good eyes can see down to about $+6$ magnitudes.

but are mixtures of many elements and isotopes, with one eponymous element. The above description is a gross simplification of the composition and nuclear astrophysics of a massive star. Curiously, although the inner regions are hot, their high densities and the ordered arrangement of their constituent nucleons into nuclei results in a low specific entropy. In fact, from birth to death, the drama of a star's evolution is the inexorable decrease of its core entropy. At low entropies, pressure is derived mostly from the zero-point Fermi motion of degenerate free electrons and not from the ideal gas of nuclei. Importantly, the concept of a limiting Chandrasekhar mass is relevant only for such degenerate material.

At core collapse, the outer shells are not moved. They are oblivious to their impending fate until the supernova shock wave generated in the core hits and ejects them. Hence, the basic composition of most of the debris is determined during the quiescent burning stages of the massive star. The supernova explosion merely lifts the shells into space (at speeds of $\sim 10,000$ to $\sim 20,000 \text{ km s}^{-1}$). However, shocked inner-mantle material does achieve temperatures sufficient to explosively burn some of the 'silicon' shell into iron-peak nuclei, some of the 'oxygen' shell into iron-peak and intermediate-mass nuclei, and some of the 'carbon' shell into oxygen. This post-processing alters the yields of the intermediate-mass and iron-group elements and is responsible for the generation of radioactive isotopes, such as ^{56}Ni , ^{57}Ni and ^{44}Ti , in the inner ejecta¹⁰. These radioisotopes are important power sources for the late-time light curve of a core-collapse supernova and have in fact been directly detected by γ -ray satellites (SN1987A, ref. 8; Cas A, ref. 11). On average, the theoretical and observational yields of ^{56}Ni , ^{57}Ni and ^{44}Ti hover around $0.07M_{\odot}$, $0.0015M_{\odot}$ and $0.0001M_{\odot}$, respectively, per core-collapse supernova, though there are individual exceptions. These average yields, galactic supernova rates, and theories of galactic evolution are consistent with the current cumulative abundances in the Solar System of their stable daughter isotopes, ^{56}Fe , ^{57}Fe and ^{44}Ca .

The origin of the elements has been a central focus of astrophysics for the last forty years¹², but equally important is the determination of the mechanism by which core-collapse supernovae explode. One might have thought that the bounce and rebound of the imploding core that creates the young shock wave and drives it with a single massive thrust would have been all that was necessary to explain both the supernova blast and the simultaneous creation of a protoneutron star¹³. However, formed at a radius of ~ 20 – 30 kilometres, the shock has to plough through many tenths of solar masses of infalling material. In the process, at the many MeV ($1 \text{ MeV} \equiv \sim 10^{10} \text{ K}$) temperatures created behind the initially strong shock, the heavy nuclei it encounters are disassociated into nucleons. This is an endothermic process. In addition, and almost uniquely to supernovae, at the high temperatures and densities that are achieved during collapse, neutrinos of all three flavours (in particular, ν_e neutrinos created by electron capture on the newly-liberated free protons) are copiously radiated. The initial neutrino luminosities can reach $\sim 10^{54} \text{ erg s}^{-1}$, equivalent to $\sim 0.5M_{\odot}c^2$ per second¹⁴. Between neutrino energy losses and nuclear breakup, the young shock has no chance; such energy and pressure penalties sap it of strength and within only 10–20 ms of its birth, it stalls between a radius of 100 and 200 km into a quasi-stationary accretion shock¹⁵. Effectively, the front stops propagating outward in radius, although it continues to propagate outward in interior mass, as infalling shells of matter with speeds of $\sim 30,000$ to $\sim 70,000 \text{ km s}^{-1}$ reach it, are shock-compressed and heated, and then settle onto the nascent protoneutron star.

Since the progenitor's outer radius is 10^6 – 10^9 km and much of the matter in the accretion-shock-bounded protoneutron star has a negative velocity, this is an unsatisfactory state of affairs. The theorist's modern quest is to determine how the accretion shock is revived and revitalized into a supernova explosion (with a healthy representation of positive velocities!). In the past 15 years, there has

been substantial progress towards solving this puzzle in the pit of the enshrouding star, quite literally a riddle wrapped in a mystery inside an enigma.

The solution seems to involve neutrinos as the drivers of explosion^{16–18}. The protoneutron star is exotic on many counts, not the least of which is that despite the notoriously weakly interacting nature of neutrinos with matter, they are partially trapped in the superdense core. Upon creation, they do not stream out immediately at the speed of light, but must diffuse to escape. As the neutrino absorption and scattering mean-free-paths in the centre are metres¹⁹ and the number of mean-free-paths from the centre to infinity is $\sim 10^4$ – 10^6 , the object is a 'neutrino star' with a surface of emission, the neutrinosphere, akin to the photosphere of the Sun. Instead of leaving the protoneutron star within milliseconds, the characteristic loss time for neutrinos is seconds. It takes this long for the trapped leptons and energy to leak out of the hot and extended protoneutron star, in order to effect its conversion into the more compact, cold neutron star²⁰. Such an evolutionary timescale was verified by the detection in two deep underground mines 13 years ago over periods of ~ 5.5 and ~ 12.5 seconds of the neutrinos from SN1987A (refs 21, 22). These observations are consistent with the emission in neutrinos of all species of $\sim (2\text{--}3) \times 10^{53}$ ergs, equal to the gravitational binding energy of a neutron star and more than 100 times the kinetic energy of the blast. During this brief pulse, the supernova neutrino luminosity can rival the total optical output of the observable universe. Hence, in an energetic sense, the optical supernova is a sideshow to the main event: the neutrino burst that attends the formation of a neutron star. In principle, neutrinos are the best direct probes of the internal dynamics of both the supernova mechanism and neutron star birth^{21–29}. As a consequence, in anticipation of the next galactic core collapse, a network of massive underground neutrino telescopes that, with luck, will collect thousands of supernova neutrino events now stands guard^{25–29}.

But how does the leakage of energy by neutrino radiation from the gradually settling inner core re-energize the shock and launch an explosion? The key lies in the fact that though energy is being radiated from the inner core and this core is becoming more and more bound with time, its accretion-shock-bounded mantle is being heated by the absorption of a fraction of the escaping neutrino flux. The region just interior to the stalled shock in which there is net heating is called the 'gain' region³⁰. This neutrino-mediated energy transfer from the core to the mantle is the essence of the mechanism, as it is currently envisioned. If there were no mass accretion from the still-collapsing outer core, the neutrino-heated mantle would be unstable enough to explode³¹. The accretion pressure tamp is all that is keeping the shock bottled up. However, since mass accretion is being fuelled by zones which are infalling from progressively further out in the star, the mass accretion rate ($\dot{M}$) onto the core is inexorably subsiding. Hence, the supernova is a race between a decaying neutrino luminosity and a subsiding $\dot{M}$. Any process that expands the gain region, maintains or boosts the neutrino luminosity, or decreases $\dot{M}$ facilitates explosion. The specific details of neutrino transfer as neutrinos decouple from matter are also important³². Therefore, the stalled shock is just biding its time, perhaps for hundreds of milliseconds to seconds, until the critical condition for explosion is achieved³³.

The behaviour in the pit during this pause is the focus of current research. Variations from progenitor to progenitor in both density and rotation profiles³⁴ are all part of the mix and should result in a spread of outcomes, final neutron star masses, ^{56}Ni yields, and explosion energies. Indeed, two supernovae, SN1994W and SN1997D, have kinetic energies of $\sim 0.5 \times 10^{51} \text{ erg}$ and ^{56}Ni yields of $\sim 2.0 \times 10^{-3}M_{\odot}$, both significantly below the norm^{35,36}. This implies that the explosion mechanism supports variations from supernova to supernova, perhaps as a function of progenitor mass, that theorists have yet to explain. Some cores may be too dense to

explode before the protoneutron star itself becomes gravitationally unstable to general relativistic collapse. A black hole would then be the inevitable and irreversible result. Whether a supernova could be launched in such a circumstance is unknown. Furthermore, a supernova might be launched in canonical fashion, only to witness the subsequent collapse of its 'launching pad' into a black hole after many seconds of cooling and neutronization or matter fall back. This type of event may depend upon the existence of as yet unverified soft phases of nuclear matter at supranuclear densities^{27,38}.

There is another important and generic characteristic of supernova explosions: symmetry breaking. Rayleigh–Taylor and Kelvin–Helmholtz fluid instabilities, akin to those that explain why water falls out of a glass, why a shaken vinaigrette mixes, why a flag flaps, and why an atomic explosion on Earth assumes the shape of a mushroom cloud, are ubiquitous in supernova explosions. Indeed, the maintenance of spherical symmetry in a context of severe density gradients, subjected to extreme and variable acceleration and gravity fields, would seem bizarre. Collapse and explosion must be quite aspherical and asymmetrical. In particular, it can be shown^{31,39,40} that the gain region interior to the stalled shock is also a region of vigorous transonic convection, driven by neutrino heating 'from below'. This is analogous to the boiling of water on a stove. Many theorists believe that such convection increases the efficiency of mantle neutrino heating, increases the size of the gain region, and, therefore (as described above), is an important factor in the mechanism of explosion. Furthermore, once the shock wave is re-energized, as it propagates down the density gradient of the outer progenitor star and encounters the carbon–helium and helium–hydrogen composition boundaries, new Rayleigh–Taylor instabilities are tripped⁴¹. Clearly, there is every theoretical reason to believe that such supernova debris should be clumpy and aspherical. In fact, optical and infrared line profile measurements⁴², the images of young supernovae and SNRs, and polarization studies⁴³ all demonstrate that these explosions are far from spheres, particularly in their inner reaches. In some cases, fast rotation will impose a characteristic axis that could lead to a bipolar explosion structure.

However, one of the most spectacular indications that the explosion that leaves behind a neutron star is asymmetric in some important way comes from the radio pulsar population. It is now known that the eccentric orbits of many binaries containing neutron stars, including neutron–star–neutron–star–pulsar systems, and the observed transverse motions of radio pulsars across the sky cannot be explained without often evoking a violent kick at birth. The speed with which the nascent neutron is rocketed out of its cradle is measured to average $\sim 450 \text{ km s}^{-1}$ (ref. 44). Some pulsars⁴⁵ reach speeds of $\sim 1,500 \text{ km s}^{-1}$, far in excess of the escape velocity ($\sim 300\text{--}600 \text{ km s}^{-1}$) from our Milky Way. Such speeds cannot be explained by the slow orbit speeds of their progenitor stars in the binary systems in which many supernovae arise. Such speeds require an asymmetry in the supernova explosion itself. There may be a bimodal distribution of kicks, with a fraction with speeds near zero and a fraction with an average speed of $\sim 600 \text{ km s}^{-1}$ (ref. 46). Be that as it may, whether the young neutron star recoils due to asymmetric mass ejection or asymmetric neutrino radiation is not yet known; it is one of the central unsolved mysteries of supernova research.

The emerging supernova– γ -ray burst connection

Gamma-ray bursts were discovered in the 1960s by the Vela constellation of satellites established to monitor nuclear treaties and explosions in the near-Earth environment. For decades, these bursts, each lasting from a fraction of a second to minutes and boasting a featureless spectrum of MeV photons, were thought to be associated with our galaxy, but poor angular resolution and the lack of optical, radio, or X-ray counterparts denied us the means to pinpoint the culprits with precision. As a consequence, almost nothing was known of their source or engine and throughout the

1970s and 1980s the ratio of theories to detected photons was uncomfortably high. In 1991, the Compton Gamma-ray Observatory was launched and due to its sensitivity was expected to see the edge of the γ -ray burst population. Since this population was thought to be in the galaxy, its angular distribution was expected to be concentrated towards the Milky Way. It was not. In fact, it was isotropically distributed on the sky, implying either a local, a galactic halo, or a cosmological distribution. The breakthrough came in 1997 when the Italian–Dutch satellite, Beppo-SAX (refs 47, 48), detected first an X-ray counterpart to GRB970228 (ref. 49), and then a few months later a counterpart to GRB970508, for which respectable angular resolution was then achievable. Given a small angular error box, the sources of the bursts were identified and quick optical follow-up was possible. That follow-up yielded a fading optical source at the position of GRB970228 (ref. 50), and then an optical spectrum for GRB970508. The spectrum contained absorption lines due to singly ionized iron and magnesium (Fe II and Mg II) at a redshift of $z = 0.835$, proving that this burst at least was of cosmological origin⁵¹. Soon, many optical and radio counterparts were detected and other high redshifts were obtained. Bursts are found to coincide with galaxies in the Hubble flow, perhaps even in their star-forming regions. The γ -ray burst itself and its $\sim t^{-2}$ power-law decay afterglow is now explained in the broad context of an extremely relativistic blast wave, with a very large initial Lorentz factor $\Gamma \gtrsim 100$, that interacts with and propagates through its local gaseous environment⁵². These are found through the Universe. Many important aspects have yet to be worked out and much remains to be understood, but several recent discoveries are starting to connect some γ -ray bursts with supernovae. After a few weeks, the optical afterglows of GRB970228 (ref. 53) and GRB980326 (ref. 59) deviated from the simple power-law decay expected in the relativistic blast model and can be interpreted as a superposition of 'classic' power-law afterglows with supernova light curves. GRB980425 actually coincided in time and direction with the peculiar type Ic supernova, SN1998bw, that exploded in a nearby galaxy⁵⁵. SN1998bw had very broad spectral features and may have been more energetic than most such supernovae⁵⁶. However, from the accumulating observational record, the association between a subset of γ -ray bursts and a subset of supernovae seems excitingly credible. But what could it all mean? Statistical arguments, existing supernova data, and a variety of physical arguments indicate that most supernovae cannot be associated with a γ -ray burst. Nevertheless, since they are at cosmological distances, if they radiate isotropically, most γ -ray bursts must be "hypernovae"^{52,57} and liberate $\sim 10^{51}$ to $\sim 10^{53}$ erg, values that are interestingly close to supernova energies, to the binding energies of neutron stars, and to the orbital energies of tight binaries of compact objects such as neutron stars and black holes, but are close to the 'available' energy of almost nothing else. If γ -ray bursts are not isotropic, but beamed, then less energy is required. However, the intrinsic γ -ray burst rate must go up correspondingly. It may be that some supernovae, through the agency of very rapid rotation, strong magnetorotational effects^{58,59}, and/or neutrino–antineutrino pair production along baryon-poor bipolar channels⁶⁰, may be able to generate a relativistic beam of e^+e^- pairs that can punch through a vestigial supernova progenitor envelope to spawn a γ -ray burst and its early afterglows. The relativistic flow is later followed by the much slower supernova explosion, still recognizable as such. But it is not yet known what tail of the core-collapse distribution could do this.

Using supernovae to do cosmology

The brightness of supernovae naturally suggests their use in surveying the Universe. If they were standard candles, a comparison between their apparent brightness and their intrinsic (or 'absolute' brightness) would yield their distance. A spectrum taken with a large-aperture telescope capable of precision measurements of dim objects, made dim by distance, would yield the spectral redshift (z)

of the supernova (actually, that of its host galaxy) in the Hubble flow of the expanding Universe. A collection of these measurements would provide redshift–distance and redshift–magnitude relations which can be used to distinguish different models of the cosmos, to determine its geometry and mass-energy content, and to help determine its ultimate fate.

Surveying standard patches of the sky over months and years to discern changes in them that are characteristic of supernova explosions, two groups of astronomers^{6,7} may recently have done just that. The Supernova Cosmology Project (SCP, ref. 6) and the High-Z Supernova Search Team^{7,61} have culled more than 80 and 50, respectively, of the rarer, though brighter, type Ia supernovae. These collections include supernovae at redshifts from ~ 0.2 to ~ 1.0 and, hence, are penetrating into the non-linear regions of the Hubble flow to probe the curvature of the Universe. A ‘Hubble diagram’ for type Ia supernovae, depicting brightness versus red-

shift, is given in Fig. 2a, taken from the SCP. The high-redshift data are those newly obtained and the low-redshift data are from the previous Calán-Tololo⁶² survey of nearby type Ia supernovae. The latter provides a reference benchmark that anchors the curve of data in Fig. 2a to allow a measurement of its curvature. This curve’s change in slope has a bearing on whether the Universe is geometrically open or closed (that is, what its curvature parameter, k , is), whether it is accelerating, and whether it has a non-zero cosmological constant (Λ , Einstein’s self-proclaimed “biggest mistake”).

As Fig. 2a suggests, what both groups find is that not only will the Universe expand forever, but that it seems to be accelerating. Such acceleration implies, in the context of general relativity and the inflationary model of the Universe, that Λ , the cosmological constant, is non-zero, and more than that, that it dominates all the matter in the Universe—dark, baryonic and neutrino—in determining the current expansion rate. In the canonical inflationary universe, Ω_M and Ω_Λ are, respectively, the matter and Λ fractions of the mass-energy content of the cosmos (see Box 1). An approximate fit to the oval in Fig. 2b is: $0.8\Omega_M - 0.6\Omega_\Lambda = -0.2 \pm 0.1(1\sigma)$. Hence, given the inflationary paradigm, Λ seems to represent $\sim 70\%$ of the mass-energy equivalent of the Universe, with most of the other $\sim 30\%$ coming from the dark matter, whatever that may be. (Other arguments suggest that baryons constitute no more than $\sim 5\%$; ref. 63.) Figure 2b shows the confidence regions in Ω_Λ versus Ω_M space, as calculated by the High-Z Supernova Search Team. A value of zero for Ω_Λ is excluded to better than 3σ , particularly under the assumption of an inflationary ‘Big Bang’. A cosmological constant is akin to a non-zero mass-energy density for the vacuum, with a negative pressure, and its possible physical origin is currently a topic of hot speculation⁶⁴. If Ω_Λ is indeed non-zero, this would be one of the most revolutionary scientific discoveries of the decade.

How firm is this startling finding? How much does it depend on the physics of type Ia supernovae? In fact, though much is known about the thermonuclear explosions of white dwarfs, much is not. We do not know whether the burning wave propagates through the white dwarf as a supersonic detonation or as a subsonic deflagration. We do not understand the details of initial ignition and

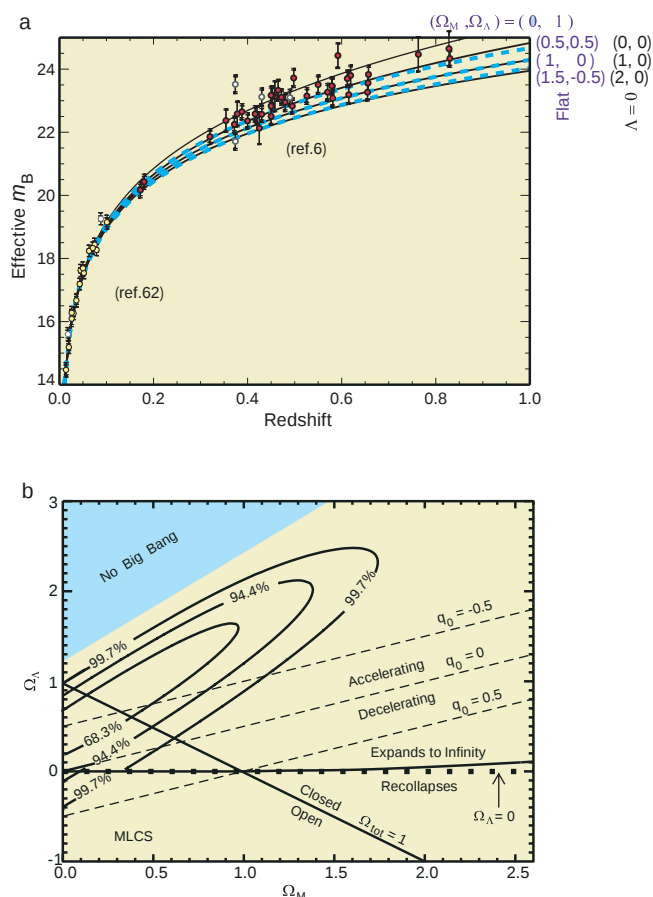


Figure 2 The effective brightness of supernovae with redshift, and how those data are used to reveal cosmological parameters. **a**, A Hubble diagram depicting the calibrated peak blue brightness L_{peak} in astronomical magnitudes ($M_B = -2.5 \log_{10} L_{\text{peak}} + \text{constant}$) versus redshift of the collection of type Ia supernovae discovered and studied by the Supernova Cosmology project (ref. 6, red and white circles). Superposed are various curves representing predictions for different cosmologies, indicated by values for $(\Omega_M, \Omega_\Lambda)$ in parentheses. Also shown are data from ref. 62 (yellow circles). The dashed curves are for $\Lambda = 0$. The data seem to fit only the uppermost curves, associated with an accelerating universe. **b**, A plot of Ω_Λ versus Ω_M (where Ω_Λ and Ω_M are the cosmological-constant and matter fractions of the mass-energy content Ω_{tot} of the cosmos) taken from the analysis done by the High-Z Supernova Team (ref. 7) using their type Ia supernova data. Shown are the 1, 2 and 3 σ confidence contours, the $\Omega_\Lambda + \Omega_M = 1$ line of the inflationary universe (curvature parameter, $k = 0$), and three dashed lines depicting solutions for various deceleration parameters (q_0). It may be seen that a solution with a non-zero Ω_Λ is favoured. (MLCS stands for Multicolor Light Curve Shapes, and the dotted curve is for $\Omega_\Lambda = 0$.)

Box 1

A brief cosmology primer

The relationship between the various cosmological parameters is

$$\begin{aligned} \left(\frac{\dot{R}}{R}\right)^2 &= \frac{8\pi}{3} G\rho - k \frac{c^2}{R^2} + \Lambda/3 \\ &= \frac{8\pi}{3} G(\rho + \Lambda') - k \frac{c^2}{R^2} \\ &= H_0^2 \left(\frac{\rho + \Lambda'}{\rho_{\text{crit}}} \right) - k \frac{c^2}{R^2} \\ &= H_0^2 (\Omega_M + \Omega_\Lambda) - k \frac{c^2}{R^2} \end{aligned}$$

where R is the scale of the universe, c is the speed of light, H_0 is the Hubble constant (the current ratio of the Hubble recession velocity to distance and a fundamental cosmological parameter), $\rho_{\text{crit}} (= 3H_0^2/8\pi G)$, is the critical mass-energy density for a given H_0 , Ω_M is the density parameter due to matter, Ω_Λ is the density parameter due to the cosmological constant, Λ' is the cosmological constant in units of mass-energy density, and ρ is the mass-energy density of matter (that for pressureless particles is proportional to $1/R^3$). The universal curvature parameter k is either 0 (flat), -1 (open), or $+1$ (closed). If the inflationary paradigm of the universe is assumed, then $\Omega_M + \Omega_\Lambda = 1$ and $k = 0$.

runaway. We do not know the nature of the binary companion that feeds the white dwarf—whether it is a main sequence star, a subgiant star, a giant star, or another white dwarf. All these types of companions might occur with some frequency, in different environments. The hint that only type Ia supernovae occur in elliptical galaxies has not clarified the situation, nor do we know whether novae, which are much weaker (by a factor of 10^6) thermonuclear explosions of hydrogen on the surfaces of accreting white dwarfs, and type Ia supernovae are connected in any way. The result of all this ignorance is that theorists cannot predict—from first principles and with sufficient precision—the ^{56}Ni yields, explosion energies, light curves, and spectra of type Ia supernovae to justify them as cosmological theodolites.

However, they need not do so. Astronomers have shown observationally that though there is variation from explosion to explosion in the peak luminosity of type Ia supernovae, there is a correlated variation in their peak duration and light curve shape. In the absence of theoretical guidance, observers have determined empirically that, for whatever reason, type Ia supernova light curves are approximately a one-parameter family that, through the use of light-curve templating⁷ and/or the empirical “Phillips” relation between peak brightness and the brightness 15 days after peak⁶⁵, can be made into standard candles. In fact, in this sense, type Ia supernovae are excellent distance indicators, providing one of the best current measures of the Hubble constant, H_0 . From the low-redshift intercept in Fig. 2a (note that this is a log-linear plot), a value of $\sim 64\text{--}69\text{ km s}^{-1}\text{ Mpc}^{-1}$ is obtained^{66,67}, amply within the error bars of competing techniques^{68,69}. However, the absence of an absolute calibration from theory means that, as with other forefront techniques for measuring H_0 , the use of supernovae is tied to the astronomical distance ladder. Unfortunately, the LMC is a rung on this ladder and the distance to the LMC is not known to better than 10%.

Nevertheless, given this H_0 , the age of the Universe can be calculated and the result is quite interesting. If the Universe were flat and $\Lambda = 0$, the age of the Universe would be a mere $2/3H_0 = 10\text{ Gyr}$. This age is incompatible with the age of the oldest globular clusters (11.5–13.5 Gyr), inferred from stellar evolution theory^{70,71}, and is also mildly at odds with the age obtained using radioactive nucleo-cosmochronometers, such as thorium-232 ($15.6 \pm 4.6\text{ Gyr}$; ref. 72), which are themselves products of core-collapse supernovae. However, an accelerating universe changes all this. With the non-zero cosmological constant, a back-extrapolation from the present epoch yields an age of $\sim 14.5\text{ Gyr}$. This age brings the various age estimates of the Universe into rough agreement, lending some credence to the idea that the Universe is indeed accelerating.

However, as with other techniques, there may be systematic effects or errors in the supernova-derived value of Λ . Neutrally absorbing dust may be scattered throughout the Universe, thereby corrupting supernova brightness measurements at large redshifts⁷³. With increasing redshift, gravitational lensing becomes a larger and larger correction. Type Ia supernovae in the early Universe and at large redshifts may have systematically different origins and/or light curves^{74,75}. This would be all the more plausible if type Ia supernovae could occur in different types of progenitor binary systems, with different evolutionary timescales and overall rates. These rates might themselves depend on galaxy type, age, and/or metal content, itself a function of age, given that supernovae of all types build up the Universe’s complement of heavy elements. The history of our galactic halo, with stars that span a full range of metal contents, may provide a key to the history of a universe in which the heavy elements, as the results of supernovae, gradually accumulated. Hence, much remains to be worked out, for supernovae are only now at the threshold of their partnership with cosmology. We need to obtain more supernovae at higher redshifts. Beyond redshifts of 1.0, much of the supernova’s light is shifted into the infrared. With

the Next Generation Space Telescope, planned for the next decade and optimized in the infrared, and with the world’s expanding array of large-aperture ground-based telescopes, we will soon capture type Ia supernovae at redshifts of 5 or 10 and core-collapse supernovae at redshifts greater than 3 (ref. 76). With such instruments, we will witness the first epochs of star and galaxy formation and the production by supernova of the first generation of heavy elements.

Thus we see that in important ways, the histories of star and galaxy formation and of supernovae are inextricably linked. Progress in understanding one demands progress in understanding the other. Today, as we attempt to fathom the mechanisms of supernova explosions, the origin of the elements, the death of stars, and the birth of neutron stars and black holes, we are simultaneously advancing the means by which we can comprehend our origins. Crucial to the development of the Universe, supernovae tell a story that goes beyond the exotic physics, the state-of-the-art numerical technique, and their role in surveying the Universe, to the heart of mankind’s ability to comprehend its home. $\square$

1. Woosley, S. E. & Weaver, T. A. The evolution and explosion of massive stars. II. Explosive hydrodynamics and nucleosynthesis. *Astrophys. J. Suppl.* **101**, 181–235 (1995).
2. Nomoto, K. & Hashimoto, M. Pre-supernova evolution of massive stars. *Phys. Rep.* **163**, 13–36 (1988).
3. Bethe, H. A., Brown, G. E., Applegate, J. & Lattimer, J. M. Equation of state in the gravitational collapse of stars. *Nucl. Phys.* **A324**, 487–533 (1979).
4. Taylor, J. H. & Cordes, J. M. Pulsar distances and the galactic distribution of free electrons. *Astrophys. J.* **411**, 674–684 (1993).
5. Nomoto, K. Evolution of 8–10 solar mass stars toward electron capture supernovae. I—Formation of electron-degenerate O + Ne + Mg cores. *Astrophys. J.* **277**, 791–805 (1984).
6. Perlmutter, S. *et al.* Measurements of Omega and Lambda from 42 high-redshift supernovae. *Astrophys. J.* **517**, 565–586 (1999).
7. Riess, A. *et al.* Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.* **116**, 1009–1038 (1998).
8. Tueller, J. *et al.* Observations of gamma-ray line profiles from SN 1987A. *Astrophys. J.* **351**, L41–L44 (1990).
9. Arnett, W. D., Bahcall, J. N., Kirshner, R. P. & Woosley, S. E. Supernova 1987A. *Annu. Rev. Astron. Astrophys.* **27**, 629–700 (1989).
10. Woosley, S. E. & Hoffman, R. D. The alpha-process and the r-process. *Astrophys. J.* **395**, 202–239 (1992).
11. Iyudin, A. F. *et al.* Emission from ^{44}Ti associated with a previously unknown galactic supernova. *Nature* **396**, 142–144 (1998).
12. Burbidge, E., Burbidge, G., Fowler, W. A. & Hoyle, F. Synthesis of the elements in stars. *Rev. Mod. Phys.* **29**, 547–650 (1957).
13. Colgate, S. A. & Johnson, H. J. Hydrodynamic origin of cosmic rays. *Phys. Rev. Lett.* **5**, 235–238 (1960).
14. Burrows, A. & Mazurek, T. J. Postshock neutrino transport and electron capture in stellar collapse. *Astrophys. J.* **259**, 330–340 (1982).
15. Bruenn, S. W. Stellar core collapse—Numerical model and infall epoch. *Astrophys. J. Suppl.* **58**, 771–841 (1985).
16. Colgate, S. A. & White, R. H. The hydrodynamic behavior of supernovae explosions. *Astrophys. J.* **143**, 626–681 (1966).
17. Bethe, H. E. & Wilson, J. R. Revival of a stalled supernova shock by neutrino heating. *Astrophys. J.* **295**, 14–23 (1985).
18. Wilson, J. R. in *Numerical Astrophysics* (eds Centrella, J. M., LeBlanc, J. M. & Bowers, R. L.) 422–435 (Jones & Bartlett, Boston, 1985).
19. Burrows, A. & Lattimer, J. M. Convection and the early evolution of neutron stars. *Phys. Rep.* **925**, 51–62 (1988).
20. Burrows, A. & Lattimer, J. M. The birth of neutron stars. *Astrophys. J.* **307**, 178–196 (1986).
21. Bionta, R. M. *et al.* Observation of a neutrino burst in coincidence with supernova 1987A in the Large Magellanic Cloud. *Phys. Rev. Lett.* **58**, 1494–1496 (1987).
22. Hirata, K. *et al.* Observation of a neutrino burst from the supernova SN1987A. *Phys. Rev. Lett.* **58**, 1490–1493 (1987).
23. Burrows, A. Supernova neutrinos. *Astrophys. J.* **334**, 891–908 (1988).
24. Totani, T., Sato, K., Dalhed, H. E. & Wilson, J. R. Future detection of supernova neutrino burst and explosion mechanism. *Astrophys. J.* **496**, 216–225 (1998).
25. Koshiba, M. The Kamiokande Detector. *Phys. Rep.* **220**, 229–231 (1992).
26. Ewan, G. T. *et al.* *Sudbury Neutrino Observatory Proposal* (Report SNO-87-12, 1987).
27. Bellotti, E. *Instrum. Meth.* **A264**, 5–9 (1988).
28. IVD Collaboration. *Nucl. Instrum. Meth.* **A264**, 5–21 (1988).
29. Icarus Collaboration. *Icarus Proposal* (LNGS Note 94/99-1, INFN preprint, GranSasso Laboratory, 1994).
30. Mayle, R. & Wilson, J. R. Supernovae from collapse of oxygen-magnesium-neon cores. *Astrophys. J.* **334**, 909–926 (1988).
31. Burrows, A., Hayes, J. & Fryxell, B. A. On the nature of core-collapse supernova explosions. *Astrophys. J.* **450**, 830–850 (1995).
32. Messer, O. E. B., Mezzacappa, A., Bruenn, S. W. & Guidry, M. W. A comparison of boltzmann and multigroup flux-limited diffusion neutrino transport during the postbounce shock reheating phase in core-collapse supernovae. *Astrophys. J.* **507**, 353–360 (1998).
33. Burrows, A. & Goshy, J. A theory of supernova explosions. *Astrophys. J.* **416**, L75–L78 (1993).
34. Fryer, C. & Heger, A. Core-collapse simulations of rotating stars. <http://xxx.lanl.gov/archive/astro-ph/9907433> (1999).

35. Turatto, M. *et al.* The peculiar type II supernova 1997D: A case for a very low ^{56}Ni mass. *Astrophys. J.* **498**, L129–L133 (1998).
36. Sollerman, J., Cumming, R. J. & Lundquist, P. A very low mass of ^{56}Ni in the ejecta of SN 1994W. *Astrophys. J.* **493**, 933–939 (1998).
37. Bethe, H. A. & Brown, G. E. A scenario for a large number of low-mass black holes in the galaxy. *Astrophys. J.* **423**, 659–664 (1995).
38. Brown, G. E., Weingartner, J. C. & Wijers, R. A. M. J. On the formation of low-mass black holes in massive binary stars. *Astrophys. J.* **463**, 297–304 (1996).
39. Herant, M., Benz, W., Hix, R., Fryer, C. & Colgate, S. A. Inside the supernova: A powerful convective engine. *Astrophys. J.* **435**, 339–361 (1994).
40. Janka, H.-T. & Müller, E. Neutrino heating, convection, and the mechanism of Type-II supernova explosions. *Astron. Astrophys.* **306**, 167–198 (1991).
41. Fryxell, B. A., Müller, E. & Arnett, D. Instabilities and clumping in SN 1987A. I—Early evolution in two dimensions. *Astrophys. J.* **367**, 619–634 (1991).
42. Wooden, D. H. *et al.* Airborne spectrophotometry of SN 1987A from 1.7 to 12.6 microns—Time history of the dust continuum and line emission. *Astrophys. J. Suppl.* **88**, 477–507 (1993).
43. Wang, L., Wheeler, J. C. & Höflich, P. Polarimetry of the type Ia supernova SN 1996X. *Astrophys. J.* **476**, L27–L30 (1997).
44. Lyne, A. & Lorimer, D. R. High birth velocities of radio pulsars. *Nature* **369**, 127–129 (1994).
45. Cordes, J. M., Romani, R. W. & Lundgren, S. C. The Guitar nebula—A bow shock from a slow-spin, high-velocity neutron star. *Nature* **362**, 133–135 (1993).
46. Fryer, C., Burrows, A. & Benz, W. Population syntheses for neutron star systems with intrinsic kicks. *Astrophys. J.* **496**, 333–351 (1998).
47. Piro, L. in *Joint European and National Astronomical Meeting, JENAM-97, 6th European and 3rd Hellenic Astronomical Conference* 284–287 (Univ. Thessaloniki, 1997).
48. Harrison, T. E. & McNamara, B. J. The optical counterparts of gamma-ray bursts. *Nature* **396**, 233–236 (1998).
49. Costa, E. *et al.* Discovery of an X-ray afterglow associated with the γ -ray burst of 28 February 1997. *Nature* **387**, 783–785 (1997).
50. van Paradijs, J. *et al.* Transient optical emission from the error box of the γ -ray burst of 28 February 1997. *Nature* **386**, 686–689 (1997).
51. Metzger, M. R. *et al.* Spectral constraints on the redshift of the optical counterpart to the γ -ray burst of 8 May 1997. *Nature* **387**, 879–881 (1997).
52. Paczynski, B. Gamma-ray burst-supernova relation. <http://xxx.lanl.gov/archive/astro-ph/9909048> (1999).
53. Reichart, D. E. GRB 970228 revisited: Evidence for a supernova in the light curve and late spectral energy distribution of the afterglow. *Astrophys. J.* **521**, L111–L115 (1999).
54. Bloom, J. S. *et al.* The unusual afterglow of the gamma-ray burst of 26 March 1998 as evidence for a supernova connection. *Nature* **401**, 453–456 (1999).
55. Galama, T. J. *et al.* An unusual supernova in the error box of the γ -ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
56. Woosley, S. E., Eastman, R. G. & Schmidt, B. P. Gamma-ray bursts and type Ic supernova SN 1998bw. *Astrophys. J.* **516**, 788–796 (1999).
57. Woosley, S. E. Gamma-ray bursts from stellar mass accretion disks around black holes. *Astrophys. J.* **405**, 273–277 (1993).
58. Wheeler, J. C., Yi, I., Höflich, P. & Wang, L. Asymmetric supernovae, pulsars, magnetars, and gamma-ray bursts. <http://xxx.lanl.gov/archive/astro-ph/9909293> (1999).
59. Ostriker, J. P. & Gunn, J. Do pulsars make supernovae? *Astrophys. J.* **164**, L95–L96 (1971).
60. MacFadyen, A. & Woosley, S. E. Collapsars: Gamma-ray bursts and explosions in “failed supernovae”. *Astrophys. J.* **524**, 262–289 (1999).
61. Schmidt, B. P. *et al.* The high-Z supernova search: Measuring deceleration and global curvature of the universe using type Ia supernovae. *Astrophys. J.* **507**, 46–63 (1998).
62. Hamuy, M. BVR light curves for 29 type Ia supernovae. *Astron. J.* **112**, 2408–2437 (1996).
63. Turner, M. S. & Tyson, J. A. Cosmology at the millennium. *Rev. Mod. Phys.* **71**, 145 (1999).
64. Steinhardt, P. J., Zlatev, I. & Wang, L. Quintessence, cosmic coincidence, and the cosmological constant. *Phys. Rev. Lett.* **82**, 896–899 (1999).
65. Phillips, M. M. The absolute magnitudes of Type Ia supernovae. *Astrophys. J.* **413**, L105–L108 (1993).
66. Suntzeff, N. *et al.* Optical light curve of the type Ia supernova 1998bu in M96 and the supernova calibration of the hubble constant. *Astron. J.* **117**, 1175–1184 (1999).
67. Gibson, B. K. *et al.* The HST key project on the extragalactic distance scale XXV. A recalibration of cepheid distances to type Ia supernovae and the value of the hubble constant. <http://xxx.lanl.gov/archive/astro-ph/9908149> (1999).
68. Kennicutt, R. C. Jr., Freedman, W. L. & Mould, J. R. Measuring the hubble constant with the Hubble space telescope. *Astron. J.* **110**, 1476–1491 (1995).
69. Sakai, S. *et al.* The Hubble space telescope key project on the extragalactic distance scale XXIV: The calibration of Tully-Fisher relations and the value of the hubble constant. <http://xxx.lanl.gov/archive/astro-ph/9909269> (1999).
70. Chaboyer, B., Demarque, P., Kernan, P. J. & Krauss, L. M. A lower limit on the age of the universe. *Science* **271**, 957–961 (1996).
71. Chaboyer, B., Demarque, P., Kewnan, P. J. & Krauss, L. M. The age of globular clusters in light of Hipparcos: Resolving the age problem? *Astrophys. J.* **494**, 96–110 (1998).
72. Cowan, J. *et al.* R-process abundances and chronometers in metal-poor stars. *Astrophys. J.* **521**, 194–205 (1999).
73. Aguirre, A. Dust versus cosmic acceleration. *Astrophys. J.* **512**, L19–L22 (1999).
74. Umeda, H., Nomoto, K., Kobayashi, C., Hachisu, I. & Kato, M. The origin of the diversity of type Ia supernovae and the environmental effects. *Astrophys. J.* **522**, L43–L47 (1999).
75. Riess, A. G., Filippenko, A. V., Li, W. & Schmidt, B. P. An indication of evolution of type Ia supernovae from their risetimes. <http://xxx.lanl.gov/archive/astro-ph/9907038> (1999).
76. Madau, P., Della Valle, M. & Panagia, N. On the evolution of the cosmic supernova rates. *Mon. Not. R. Astron. Soc.* **297**, L17–L22 (1998).
77. Aschenbach, B. Discovery of a young nearby supernova remnant. *Nature* **396**, 141–142 (1998).
78. Ashworth, W. B. A probably Flamsteed observation of the cassiopeia-A supernova. *J. Hist. Astron.* **11**, 1 (1980).

Acknowledgements

I thank A. Riess, B. Schmidt, S. Perlmutter, P. Pinto, T. Thompson, A. MacFadyen, C. Fryer and D. Frail for conversations.

Correspondence and requests for materials should be addressed to the author (e-mail: aburrows@as.arizona.edu).

Genetic ablation reveals that the roof plate is essential for dorsal interneuron specification

Kevin J. Lee*, Paula Dietrich† & Thomas M. Jessell*

* Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Centre for Neurobiology and Behavior, and † Department of Genetics and Development, Columbia University, New York, New York 10032, USA

During neural development in vertebrates, a spatially ordered array of neurons is generated in response to inductive signals derived from localized organizing centres. One organizing centre that has been proposed to have a role in the control of neural patterning is the roof plate. To define the contribution of signals derived from the roof plate to the specification of neuronal cell types in the dorsal neural tube, we devised a genetic strategy to ablate the roof plate selectively in mouse embryos. Embryos without a roof plate lack all the interneuron subtypes that are normally generated in the dorsal third of the neural tube. Using a genetically based lineage analysis and *in vitro* assays, we show that the loss of these neurons results from the elimination of non-autonomous signals provided by the roof plate. These results reveal that the roof plate is essential for specifying multiple classes of neurons in the mammalian central nervous system.

The diversification of cell types in vertebrate embryos depends on inductive signals produced by specialized cell groups that act as organizing centres^{1,2}. In some cases, primary organizing centres induce the formation of secondary signalling centres that secrete the same or related signalling molecules^{3–5}. For example, at the midline of the neural plate, signals from the notochord mediated by Sonic hedgehog (Shh) direct the formation of floor-plate cells, which also express Shh⁶. It remains unclear, however, whether the signalling functions of sequential organizers such as the notochord and floor plate have distinct functions, or whether these inductive cascades simply prolong the availability of a single, multifunctional signal.

In the dorsal neural tube, cell specification appears to involve a similar inductive cascade, mediated by signals provided initially by the epidermal ectoderm and later by the roof plate⁷. The epidermal ectoderm and the roof plate are each sufficient to induce neural crest cells and distinct classes of dorsal interneurons *in vitro*^{8–10}. Both epidermal ectoderm and roof-plate cells express several members of the bone morphogenetic protein (BMP) family, each of which mimics the ability of the epidermal ectoderm and roof plate to induce these dorsal neural cell types *in vitro*. Moreover, inhibition of BMP signalling blocks the ability of both the epidermal ectoderm and the roof plate to induce dorsal neural cell types^{9,10}.

Although these results show that BMP signalling is involved in the specification of dorsal neural cells, they do not resolve whether the BMP signals derived from the epidermal ectoderm and the roof plate have distinct or redundant functions in patterning the dorsal neural tube. Here we have addressed this issue by examining the generation of dorsal neural cell types in mice in which the roof plate has been selectively ablated by toxin gene expression. Although the signalling activities of the epidermal ectoderm appear to be intact, the loss of the roof plate in these embryos is associated with failure to generate virtually all of the neuronal classes that are normally generated in the dorsal region of the neural tube. Thus, our results establish that the epidermal ectoderm and roof plate have distinct functions in the control of the dorsal neural pattern. More generally, they indicate that cascades of inductive signalling may involve sequentially acting organizers, each with a specialized role in the promotion of cell diversity during vertebrate embryogenesis.

A genetic strategy for roof-plate ablation

The respective contributions of the roof plate and the epidermal

ectoderm to dorsal neuronal patterning have not been defined by conventional surgical ablation methods. We therefore developed a genetic method to eliminate the roof plate in mouse embryos by directing selective expression of a cellular toxin gene^{11–13} and examined the effect of roof-plate loss on neuronal differentiation

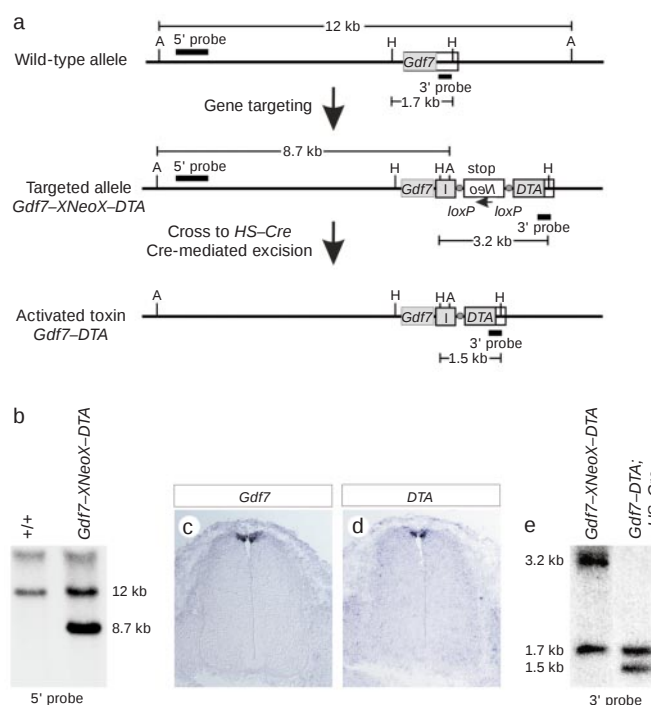


Figure 1 Conditional expression of a diphtheria toxin gene in roof-plate cells. **a**, Targeting strategy (see Methods). **b**, Southern blot of wild-type and *Gdf7-XNeoX-DTA/+* embryonic stem cell DNA, digested with *Asp718* (A) and probed with the 5' probe, showing the wild-type (12 kb) and recombinant (8.7 kb) alleles. **c**, **d**, Expression of *Gdf7* (**c**) and the inactive *Gdf7-XNeoX-DTA* gene (**d**) is restricted to roof-plate cells. **e**, Southern blot of yolk sac DNA from *Gdf7-XNeoX-DTA/+* and *Gdf7-DTA/+; HS-Cre/+* embryos digested with *HindIII* and probed with the 3' probe, showing loss of the inactive toxin allele (3.2 kb) and appearance of the activated *Gdf7-DTA* allele (1.5 kb).

in the dorsal neural tube. The gene encoding the diphtheria toxin A subunit (DTA)¹⁴ was introduced into the *Gdf7* locus, which in neural tissue is restricted to the roof plate, by homologous recombination in embryonic stem cells. To prevent the expression of the toxin gene during propagation of the mouse strain, we used the Cre recombinase-loxP system^{15,16} to regulate DTA expression in a conditional manner (Fig. 1a). We constructed a Cre-regulated DTA cassette (*XneoX-DTA*; see Fig. 1), in which translation of the DTA open reading frame is initially prevented by interposition of a sequence rich in termination codons, the reversed *Pgk-neo* gene, flanked by loxP sites. Cre-mediated recombination between the loxP sites removes this stop sequence and restores the DTA reading frame, thus permitting DTA expression. Mice heterozygous for the unrecombined *Gdf7-XneoX-DTA* allele were viable and fertile. In such mice, expression of the *Gdf7-XneoX-DTA* gene was excluded from the epidermal ectoderm and confined to the roof plate, a pattern that recapitulates the profile of expression of *Gdf7* (ref. 10) (Fig. 1c, d).

To generate embryos expressing the activated *Gdf7-DTA* allele, we crossed heterozygous *Gdf7-XneoX-DTA* mice with a strain of *Hs-cre* mice that express Cre transiently at the 1–2-cell stage^{17,18}. In 64% (18/28) of embryos heterozygous for both the *Hs-cre* and *Gdf7-XneoX-DTA* alleles, we detected only the recombined *Gdf7-DTA* allele (Fig. 1e), indicating that all embryonic tissues had inherited the active *Gdf7-DTA* locus. As a consequence, the functional *DTA* gene was expressed selectively by roof-plate cells, and these embryos showed profound abnormalities in the development of the central nervous system (Fig. 2a, e). Forebrain and midbrain structures were markedly reduced in size and the neural tube failed to close at midbrain and hindbrain levels (Fig. 2e). In contrast, the

neural tube was closed at spinal levels but was noticeably smaller than normal (Fig. 2 f–h), especially in the most caudal regions (data not shown). The overall development of non-neural structures was not obviously affected and *Gdf7-DTA* embryos remained viable until at least embryonic day (E)12.5.

Roof-plate deletion in *Gdf7-DTA* embryos

To examine the timing and selectivity of ablation of dorsal midline neural cells in *Gdf7-DTA* embryos, we monitored the expression of genes that are normally restricted to cells at the dorsal midline of the neural tube, or in the overlying epidermal ectoderm. *Bmp2* is expressed by epidermal ectodermal cells overlying the dorsal neural tube^{19,20} and its expression pattern was unchanged in caudal regions of E9.5 *Gdf7-DTA* embryos (Fig. 2b, f). Moreover, as described below, the inductive signalling properties of epidermal ectoderm are not perturbed in *Gdf7-DTA* embryos (Fig. 2i–l). *Wnt1*, *Msx1* and *Zic2* are expressed during normal embryogenesis by cells in the dorsal neural tube before the onset of *Gdf7* expression^{21–24}; the expression of these three genes by cells at the dorsal midline of the caudal neural tube was maintained in E9.5 *Gdf7-DTA* embryos (Fig. 2c, d, g, h; and data not shown). Thus, the initial steps in the specification of dorsal midline cells occur normally in *Gdf7-DTA* embryos. Neural crest cells are among the earliest cell types to be specified at the dorsal midline of the neural tube and their differentiation appears to be induced by signals from the adjacent epidermal ectoderm^{7,8,25,26}. Neural crest cells were generated normally in *Gdf7-DTA* embryos, as assessed by expression of *Hfh2* and *Ng2* (refs 27 and 28; and data not shown) and by the appearance of dorsal root ganglia and other differentiated neural crest derivatives (Fig. 4e, k; and data not shown).

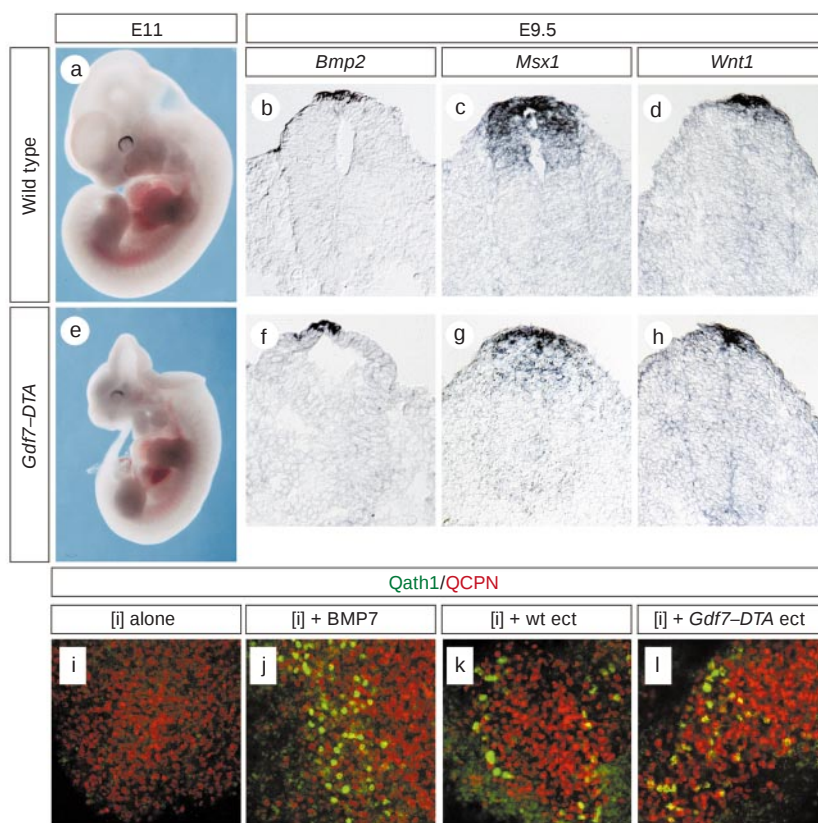


Figure 2 Initial steps of dorsal neural patterning occur normally in *Gdf7-DTA* embryos. At E11, *Gdf7-DTA* embryos (e) display an open anterior neural tube and an underdeveloped midbrain and hindbrain compared to wild-type embryos (a). Gene expression in the E9.5 wild-type (b–d) and *Gdf7-DTA* (f–h) dorsal neural tube. *Bmp2* expression (b, f) in ectodermal cells overlying the dorsal midline of the caudal neural tube, and dorsal neural

expression of *Msx1* (c, g) and *Wnt1* (d, h), persist in *Gdf7-DTA* embryos. i–l, Epidermal ectoderm isolated from E9.5 wild-type (k) or *Gdf7-DTA* embryos (l) induces dorsal neural progenitors that express Qath1 (the quail homologue of Math1, green) in quail intermediate [i] region neural plate explants (stained with QCPN, red).

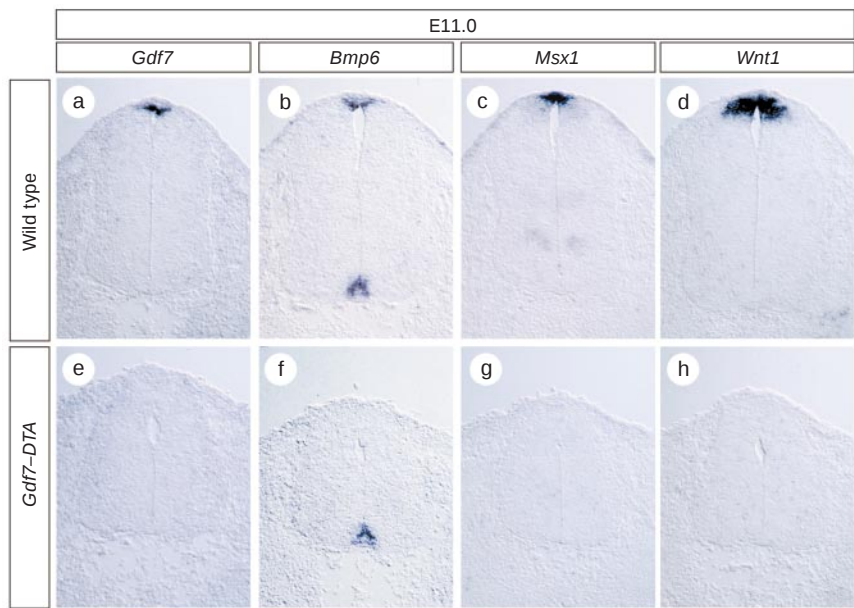


Figure 3 *Gdf7*–*DTA* embryos lack roof-plate cells. In wild-type embryos examined at E11.5 (a–d), roof-plate cells express *Gdf7* (a), *Bmp6* (b), *Msx1* (c) and *Wnt1* (d). In E11.5 *Gdf7*–*DTA* embryos (e–h), dorsal expression of these genes is eliminated.

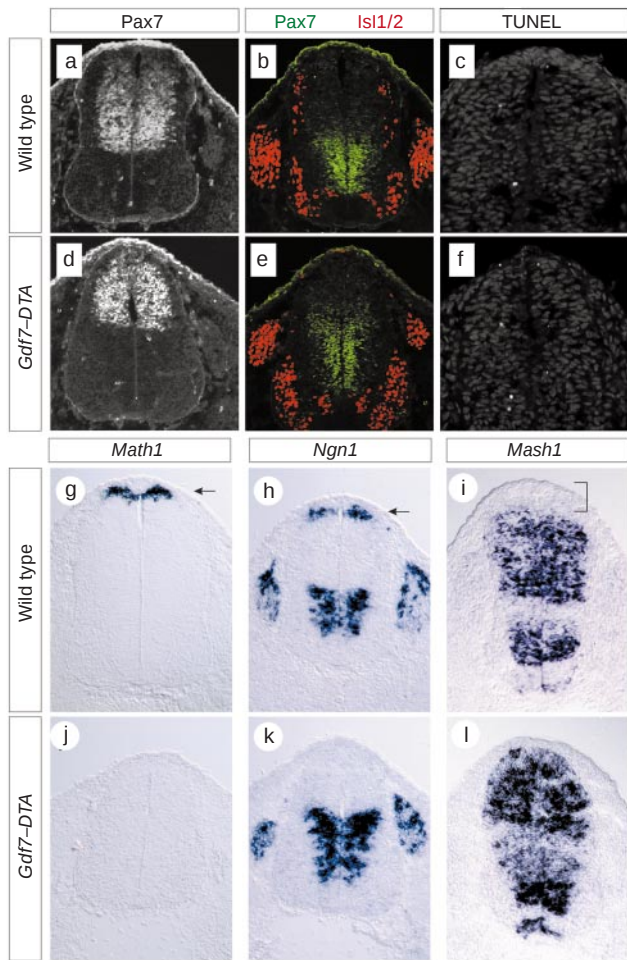


Figure 4 Loss of dorsal neural progenitors in *Gdf7*–*DTA* embryos. At E11, Pax7 is expressed by dorsal neural progenitors in both wild-type (a) and *Gdf7*–*DTA* (d) embryos. Patterning and differentiation of ventral neurons and dorsal root ganglia appear normal in *Gdf7*–*DTA* embryos, as revealed by Isl1/2 and Pax6 expression (b, e). TUNEL analysis of E10.5 dorsal neural tube reveals few apoptotic cells in both wild-type (c) and *Gdf7*–*DTA* embryos (f). However, dorsal neural progenitors that express *Math1* (arrow in g) and *Neurogenin1* (*Ngn1*) (arrow in h) are absent in *Gdf7*–*DTA* embryos (j, k). *Mash1*⁺ progenitors, normally excluded from the dorsal neural tube (square bracket in i), occupy the dorsal midline in *Gdf7*–*DTA* embryos (l).

The onset of expression of *Gdf7* by cells at the dorsal midline of the spinal cord occurs between E9.5 and E10 and is accompanied by the expression of a set of genes that characterizes definitive roof-plate cells. Both *Bmp6* and *Bmp7* are expressed by roof-plate cells following closure of the neural tube^{10,19,20,29,30}. Neither *Bmp6* nor *Bmp7* expression was detected in the dorsal tube of *Gdf7*-DTA embryos at any stage up to E12.5 (Fig. 3f; and data not shown). In addition, between E9.5 and E11, the expression of *Msx1*, *Zic2*, *Wnt1* and *Wnt3a* becomes restricted to cells in and immediately adjacent to the roof plate^{21–24,31}. By E11, none of these genes was expressed at the dorsal midline of the neural tube in *Gdf7*-DTA embryos (Fig. 3g, h; and data not shown). Moreover, cells at the dorsal midline of the neural tube in *Gdf7*-DTA mice did not exhibit the characteristic morphology of roof-plate cells³² (Fig. 3e–h; and data not shown). Together, these findings establish that *Gdf7*-DTA embryos lack roof-plate cells.

Progenitor cell defects after loss of roof plate

To assess whether loss of the roof plate influences neural patterning, we first examined the expression of Pax transcription factors that reveal the early subdivision of the neural tube into dorsal and ventral halves. The expression of Pax7 defines all dorsal neural progenitors but is excluded from ventral progenitor cells^{8,33}. In *Gdf7*-DTA embryos, the dorsal restriction in expression of Pax7 was maintained (Fig. 4a,d), showing that the establishment of generic dorsal and ventral progenitor cell identities is not affected by loss of the roof plate. However, the extent of the dorsal (Pax7⁺) progenitor domain was reduced in *Gdf7*-DTA embryos, markedly so at more caudal levels of the neural tube (data not shown). Wnt proteins derived from the dorsal midline appear to be essential signals that promote the proliferation of dorsal neural progenitors^{34,35} and thus the decrease in size of the dorsal progenitor domain may result from the elimination of *Wnt* gene expression by dorsal midline cells in *Gdf7*-DTA embryos (Fig. 3h). The ventral progenitor domain, in contrast, appeared to be slightly expanded in *Gdf7*-DTA embryos, as revealed by the extent of the domain of ventral Pax7⁺ and Pax6⁺ cells (Fig. 4d, e) that gives rise to motor neurons and ventral interneurons³⁶.

Next we examined the impact of roof-plate ablation on the specification of progenitor cell subclasses within the dorsal neural tube, defined by their expression of basic-helix–loop–helix (bHLH) transcription factors. Neural progenitors flanking the roof plate express *Math1* (refs 10, 37 and 38), whereas adjacent, more ventrally located progenitors express *Neurogenin1* (*Ngn1*) (refs 10 and 39). *Math1*⁺ and *Ngn1*⁺ progenitors were absent from the dorsal spinal cord of *Gdf7*-DTA embryos at all developmental stages examined (Fig. 4j, k; and data not shown). Instead, the most dorsal neural progenitors in *Gdf7*-DTA mice expressed *Mash1* (Fig. 4l), a gene

that is normally expressed by progenitor cells in the intermediate region of the neural tube⁴⁰. The loss of dorsal *Math1*⁺ and *Ngn1*⁺ progenitors does not appear to result from increased apoptosis in the dorsal neural tube, as shown by the similarly low incidence of TUNEL⁺ cells in wild-type and *Gdf7*-DTA embryos examined at E9.5 and E10.5 (Fig. 4c, f; and data not shown).

We also examined whether the loss of dorsal neural cell types in *Gdf7*-DTA embryos might result from a secondary change in the inductive activity of the epidermal ectoderm as a consequence of roof-plate ablation. To test this, we compared the ability of epidermal ectoderm isolated from E9.5 wild-type and *Gdf7*-DTA embryos to induce dorsal cell types in explants of the intermediate region of the neural plate dissected from stage 10 quail embryos. When cultured alone, such explants do not give rise to dorsal neural progenitors expressing *Qath1*, the quail homologue of *Math1* (Fig. 2i). In contrast, intermediate explants cultured with either wild-type or *Gdf7*-DTA epidermal ectoderm generated similar numbers of *Qath1*⁺ progenitors (Fig. 2k, l). These results provide functional evidence that the signalling properties of the epidermal ectoderm flanking the dorsal neural tube are intact in *Gdf7*-DTA embryos.

Neuronal patterning after roof-plate ablation

The loss of dorsal neural progenitors in *Gdf7*-DTA embryos was accompanied by the absence of specific classes of dorsal sensory interneurons. The *Math1*⁺ progenitor population gives rise to both D1A and D1B interneurons, which express, respectively, the LIM homeodomain transcription factors LH2A (*Lhx2*) and LH2B (*Lhx9*) (refs 10, 38). Neither LH2A nor LH2B were expressed in the dorsal spinal cord of *Gdf7*-DTA embryos at any of the developmental stages examined (Fig. 5e), indicating that the generation of both D1A and D1B neurons fails in embryos lacking a roof plate. D2 interneurons, marked by expression of *Isl1* (ref. 9), were also missing in *Gdf7*-DTA embryos (Fig. 5f).

The loss of these three populations of dorsal interneurons in *Gdf7*-DTA mice was accompanied by a marked dorsal shift in the domain of generation of a distinct set of dorsal (D3) interneurons that express *Lim1/Lim2* but not *Isl1*. Indeed, in *Gdf7*-DTA embryos, D3 neurons were detected at the dorsal midline of the spinal cord in the position normally occupied by the roof plate (Fig. 5g). A set of dorsal interneurons that express *Lmx1b* was also found in a more dorsal position in *Gdf7*-DTA embryos than in wild-type littermates (data not shown). In contrast to the profound alteration in dorsal neuronal patterning, motor neurons and ventral interneurons were generated in normal or slightly increased numbers in the ventral spinal cord of *Gdf7*-DTA embryos (Fig. 4k; and data not shown).

Thus, the analysis of bHLH and LIM-homeobox gene expression

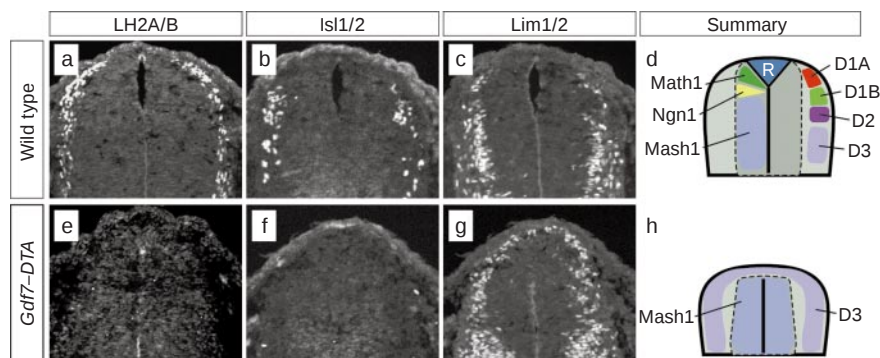


Figure 5 Absence of dorsal interneuron classes in *Gdf7*-DTA embryos. Differentiation of neuronal subclasses in the dorsal neural tube in E11 wild-type (a–d) and *Gdf7*-DTA embryos (e–h). D1A and D1B neurons expressing LH2A and LH2B (a) and D2 neurons expressing *Isl1* (b) are eliminated in *Gdf7*-DTA mice (e, f). D3 neurons expressing

Lim1/2, normally generated from a more ventral position (c), are found throughout the dorsal third of the neural tube, including the dorsal midline, in *Gdf7*-DTA embryos (g). d, h, Summary of the dorsal-to-ventral shift in progenitor cell identity and neuronal fates in *Gdf7*-DTA mice.

indicates that ablation of the roof plate leads to a failure in the generation of at least three classes of dorsal interneurons. Concomitantly, neuronal progenitors and neurons characteristic of a more ventral position come to occupy a dorsal position (Fig. 5d, h). Thus, dorsal neuronal subtypes can be divided into two major classes on the basis of their dependence on roof-plate signalling. D1A, D1B and D2 neurons are members of a roof-plate-dependent class, whereas D3 and *Lmx1b*⁺ neurons are generated independently of the roof plate.

Tracing the descendants of roof-plate cells

Lineage-tracing studies in chick have revealed that some dorsal interneurons derive from precursor cells that originate at the dorsal midline of the neural tube^{26,41}. This finding raises the possibility that the loss of dorsal interneurons in *Gdf7*–*DTA* embryos results from the cell-autonomous ablation of their dorsal midline progenitors, as a consequence of transient expression of the *Gdf7*–*DTA* gene. To address this issue, we investigated whether any of the three classes of dorsal interneurons that are missing in *Gdf7*–*DTA* mice derive from dorsal midline neural cells that have expressed *Gdf7* at an earlier point in their lineage. We devised a Cre recombinase-based method^{42–44} for tracing the lineage of the dorsal midline cells that express *Gdf7*. We used homologous gene targeting in embryonic stem cells to generate a mouse strain in which Cre recombinase expression is directed by *Gdf7* regulatory elements (Fig. 6a). The expression of Cre in these *Gdf7*–*cre* mice faithfully recapitulated expression of the endogenous *Gdf7* gene (Fig. 6b, c).

To trace the fate of cells that derive from progenitors at the dorsal midline that express *Gdf7*, we crossed *Gdf7*–*cre* mice to two independent reporter strains (ROSA26 and *cACT*) in which persistent expression of LacZ is activated by a Cre-mediated recombination event^{43,44}. In embryos carrying the *Gdf7*–*cre* allele and the ROSA26 reporter⁴⁴, LacZ was expressed by roof-plate cells, neural crest-derived neurons and scattered dorsal interneurons (Fig. 6d). The cytoplasmic expression of LacZ in these cells, however, made it difficult to identify unambiguously the neurons that derived from *Gdf7*–expressing midline cells. We therefore analysed *Gdf7*–*cre* embryos carrying the *cACT* reporter element⁴³ that directs nuclear expression of LacZ (nLacZ) in postmitotic neurons. In embryos heterozygous for both the *Gdf7*–*cre* allele and the *cACT* reporter, fewer than 0.4% of LH2A⁺ D1A neurons (Fig. 6g) or Isl1⁺ D2 neurons (Fig. 6h–j) expressed nLacZ, but ~25% of LH2B⁺ D1B neurons were labelled (Fig. 6k–m). The absence of a significant lineage relationship between *Gdf7*–expressing midline cells and D1A and D2 neurons indicates that the loss of these neuronal populations in *Gdf7*–*DTA* mice results from the elimination of roof-plate-derived signals. In the case of D1A neurons, genetic studies have shown that this signal is mediated by GDF7 (ref. 10).

Rescue of dorsal interneuron generation *in vitro*

These lineage-tracing experiments indicated, however, that ~25% of D1B neurons arise from precursors that express *Gdf7*, indicating that elimination of roof-plate cells might result in a cell-autonomous loss of some D1B neurons. Moreover, the potency of DTA⁴⁵ raised the concern that the loss of all D1B neurons *in vivo* might occur through low level expression of the *Gdf7*–*DTA* gene in D1B neurons themselves, or in their Math1⁺ progenitors. We reasoned that the ability of roof-plate signals to rescue D1B neurons in *Gdf7*–*DTA* neural tissue would argue against the idea that their loss in *Gdf7*–*DTA* mice is the consequence of the autonomous ablation of all cells in the D1B lineage. We therefore tested whether Math1⁺ progenitors and LH2B⁺ D1B neurons can be generated in the dorsal region of the neural tube in *Gdf7*–*DTA* embryos cultured in contact with a wild-type roof plate. *Gdf7*–*DTA*-derived neural tube explants grown alone did not produce Math1⁺ progenitors and LH2B⁺ D1B neurons (Fig. 7b, e). In contrast, when quail roof-plate tissue was placed next to the dorsal region of neural tube explants isolated

from *Gdf7*–*DTA* embryos, many Math1⁺ progenitors were generated (Fig. 7c). Similarly, many LH2B⁺ D1B neurons were detected in dorsal neural tube tissue derived from *Gdf7*–*DTA* embryos grown in contact with roof-plate cells (Fig. 7f).

The ability of roof-plate tissue to rescue D1B neurons indicates that the elimination of the majority of D1B neurons does not reflect low-level DTA expression in these neurons. Thus, their loss in *Gdf7*–*DTA* embryos also appears to result from the elimination of a secreted signal from the roof plate. Moreover, as *Gdf7*–*DTA* neural tissue can generate Math1⁺ progenitors and D1B neurons when provided with roof-plate-derived signals, their absence *in vivo*

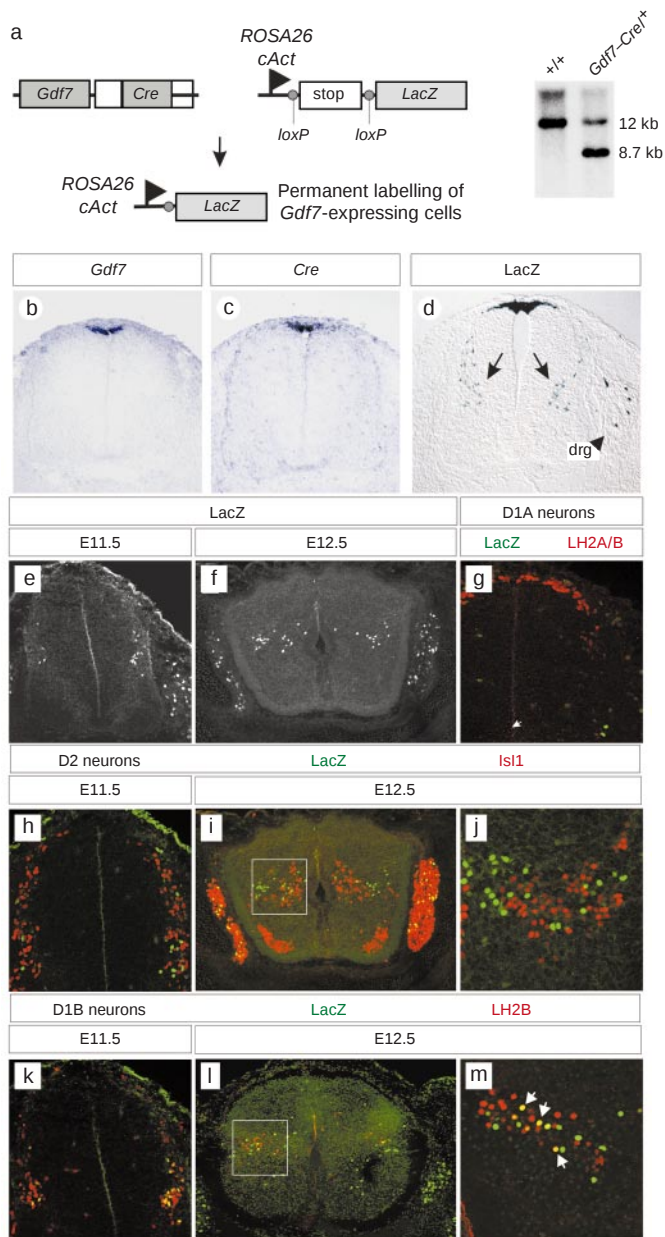


Figure 6 Lineage relationship between dorsal midline cells and D1B neurons. **a**, Lineage-tracing approach (see Methods). **b**, **c**, Expression of Cre (**c**) in *Gdf7*–*cre* mice recapitulates the roof-plate-specific expression of *Gdf7* (**b**). **d**, In *Gdf7*–*cre*; *ROSA26*–*LacZ* reporter embryos, cytoplasmic LacZ is detected in the roof plate, DRG and a subset of spinal neurons (arrows). **e**–**m**, In *Gdf7*–*cre*; *cAct*–*nLacZ* reporter embryos (**e**, **f**), nLacZ is detected in scattered DRG and spinal neurons. **g**, LH2A⁺ D1A neurons, located adjacent to the roof plate at E11.5, are distant from nLacZ⁺ cells (arrowhead). **h**–**j**, nLacZ is detected in few (0.4%) Isl1⁺ D2 neurons. **j**, Higher magnification image (boxed in **i**). **k**–**m**, nLacZ expression is detected in ~25% of LH2B⁺ D1B neurons. **m**, Higher magnification image (boxed in **l**).

establishes that no other signalling centre compensates for loss of the roof plate in the specification of these cells. Together, these results indicate that secreted signals from the roof plate are required for the generation of D1A, D1B and D2 neurons.

Distinct roles for ectoderm and roof plate

Previous *in vitro* studies using BMP antagonists have indicated that the neuronal inductive activities of both the epidermal ectoderm and roof plate are mediated by BMPs^{9,10}, but did not resolve whether BMPs derived from the epidermal ectoderm or roof plate are responsible for the induction of dorsal interneurons. Our results show that the roof plate is the sole source of signals involved in the specification of many dorsal interneurons. The failure of epidermal ectoderm signalling to compensate for the loss of roof-plate cells may result from the separation of the dorsal neural tube from the overlying epidermal ectoderm soon after neural tube closure. In addition, the presence of mesenchymal cells interposed between the epidermal ectoderm and the dorsal neural tube could provide a cellular barrier that prevents epidermally derived BMP signals from reaching the dorsal neural tube.

A loss of differentiated roof-plate cells has also been observed in mouse embryos homozygous for the *dreher* (*dr*) mutation, which results from a loss-of-function mutation in the LIM-homeobox gene *Lmx1a* (ref. 46). In *dr* mutant embryos, Math1⁺ neural progenitors and D1 neurons are generated within the dorsal neural tube, albeit in reduced numbers. Thus, dorsal neural patterning appears to be less severely affected in *dr* mutants than in *Gdf7*–*DTA* embryos, in which a complete loss of Math1⁺ progenitors and D1 neurons is observed. Although many roof-plate markers appear to be absent in *dr* mutants, residual roof-plate signalling activity could persist at levels sufficient to induce Math1⁺ cells and D1 neurons. *Lmx1a* is, however, also expressed in cells adjacent to the dorsal neural tube, and axial vertebral structures are missing in *dr*

mutants⁴⁶. These observations raise the additional possibility that *dr* embryos lack the mesenchymal tissue layer that is normally sandwiched between the epidermal ectoderm and the dorsal neural tube. If so, an apposition of epidermal ectoderm and dorsal neural tube in *dr* mutants could compensate for the absence of the roof plate, partially restoring dorsal cell-type specification.

Cell patterning in the other regions of the neural tube also appears to depend on the actions of inductive signals provided by sequentially generated organizing centres^{5,47}. In the ventral neural tube, two such organizing centres, the notochord and floor plate, can specify motor neurons and ventral interneurons⁴⁷. However, signals provided by the notochord and floor plate have been suggested to act in a redundant manner to pattern ventral neural cell types^{48–50}. In contrast, the epidermal ectoderm and roof plate clearly possess non-redundant functions in the specification of dorsal neural cell types. More generally, our findings indicate that other sequentially generated organizers may have distinct roles in directing cell identity and pattern in the vertebrate embryo. □

Methods

Gene targeting and mouse breeding

We generated the *Gdf7*–*XNeoX*–*DTA* targeting construct by inserting a 3-kilobase (kb) cassette consisting of an internal ribosome entry site (IRES, I), followed by a reversed *Pgk*–*neo* (*Neo*) gene flanked by *loxP* sites, and the *DTA* coding sequence, into a *Gdf7* genomic clone¹⁰ at an *SphI* site located immediately 3' to the *Gdf7* coding region. The *Gdf7*–*cre* targeting construct was generated by inserting a 3.4-kb *IRES*–*cre*–*FRTNeoFRT* cassette (a gift from A. Pierani) at the same position. Targeting constructs were electroporated into W9.5 embryonic stem (ES) cells and recombinants were identified by Southern blotting as shown in Figs 1b and 6a, respectively. ES cell DNA was digested with *Asp718* and probed with the 5' probe indicated in Fig. 1a to reveal a 12-kb wild-type band and an 8.7-kb band indicative of the recombinant allele. Chimaeric founder mice that transmitted the targeted alleles were produced by blastocyst injection. *Gdf7*–*DTA* embryos were generated by mating heterozygous *Gdf7*–*XNeoX*–*DTA* females to homozygous and hemizygous *Hs*–*cre6* males¹⁷. For lineage analysis, *Gdf7*–*cre* heterozygotes were mated to heterozygous *ROSA26*–*LacZ* reporter mice⁴⁴ or heterozygous *cAct*–*nLacZ* reporter mice⁴⁵. We confirmed that the *c-Act* promoter is active in all dorsal interneuron classes by

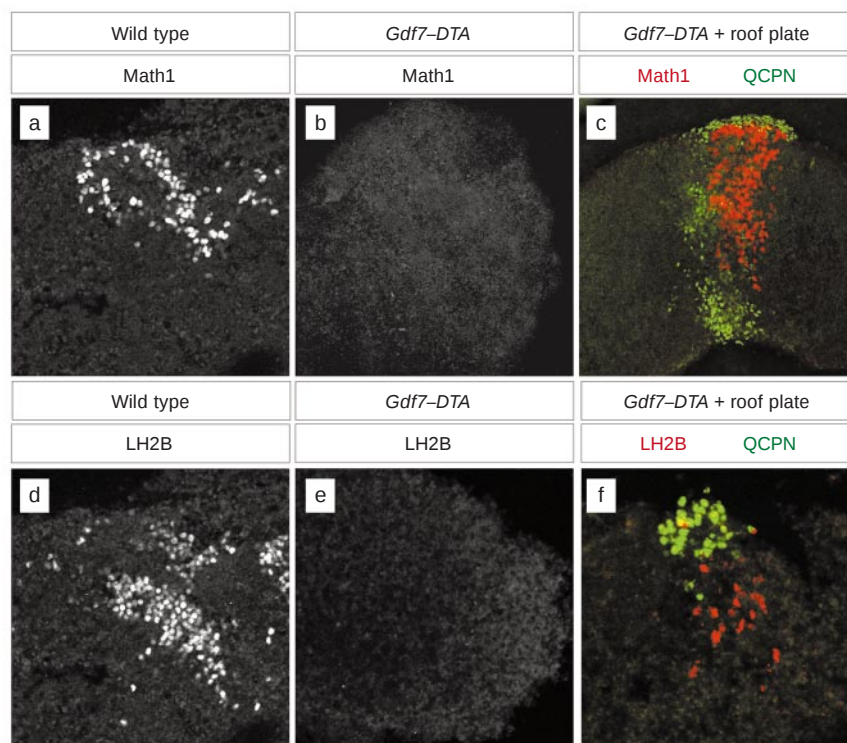


Figure 7 Roof-plate rescue of dorsal neural progenitors and dorsal interneurons in *Gdf7*–*DTA* neural tissue. **a**, Wild-type neural tube explants generate numerous Math1⁺ D1 progenitors after 48 h in culture. **b**, Neural tube explants from *Gdf7*–*DTA* embryos lack Math1⁺ D1 progenitors. **c**, *Gdf7*–*DTA* neural explants cultured in contact with quail roof

plate (stained with QCPN, green) generate many Math1⁺ progenitors (red). **d**, Wild-type neural explants cultured for 48 h generate LH2B⁺ D1B neurons. **e**, *Gdf7*–*DTA* explants do not give rise to D1B neurons. **f**, In *Gdf7*–*DTA* neural explants cultured with quail roof plate, numerous LH2B⁺ D1B neurons (red) are generated.

examining nLacZ expression in *Hs-cre*; *cAct-nLacZ* (data not shown) and *Isl1-cre*; *cAct-nLacZ* embryos (C.William, K.J.L. and T.M.J., unpublished data).

Immunocytochemistry and *in situ* hybridization

Protein expression and apoptotic cell death were detected as described¹⁰ using antisera to Pax7, Pax6, Isl1/2, Lim1/2, LH2A/B, LH2B and Math1/Qath1 (refs 9, 10, 36, 38). Rabbit anti-Brn3.0 serum was provided by E. Turner. LacZ was detected using goat anti- β -galactosidase (Arnel) or rabbit anti- β -galactosidase (Cappel), or by activity staining with X-gal. Quail cells were localized with the monoclonal antibody QCPN (Developmental Studies Hybridoma Bank). *In situ* hybridization was performed as described¹⁰ using probes to *Gdf7*, *Msx1*, *Wnt1*, *Bmp6*, *Math1*, *Neurogenin1* and *Lmx1b*. Other probes were *Mash1* (ref. 40), *Neurogenin2* (ref. 28) and *Hfh2* (ref. 27). The *Gdf7-XNeoX-DTA* messenger RNA was detected using a 2.4-kb probe comprising the DTA coding region, the *Pgk-neo* sequence and the IRES. The *Gdf7-cre* mRNA was detected with a 0.7-kb probe internal to the *Cre* coding sequence.

Explant cultures

Neural tube tissue and flanking epidermal ectoderm from wild-type and *Gdf7-DTA* mouse embryos (12–18 somites), intermediate region ([i]) neural plate from Hamburger–Hamilton (HH) stage 10 quail embryos and roof plate from HH stage 24 quail were dissected in L15 medium, dissociated in 1 mg ml⁻¹ dispase and cultured in collagen matrices at 37 °C. Mouse neural tube explants were cultured in isolation or with quail roof-plate tissue placed in contact with the dorsal region, in DMEM/F12 (1:1) (Speciality Media) supplemented with 5% rat serum (Gibco). Quail neural plate explants were cultured in isolation or with mouse ectoderm in Iscove's modified Dulbecco's medium (Gibco) supplemented with Mito serum extender (Collaborative Research). Human BMP7 and mouse GDF7 were obtained by transfection of COS-7 cells as described¹⁰.

Received 13 October; accepted 22 December 1999.

1. Harland, R. & Gerhart, J. Formation and function of Spemann's organizer. *Annu. Rev. Cell. Dev. Biol.* **13**, 611–667 (1997).
2. Pearse, R. V. & Tabin, C. J. The molecular ZPA. *J. Exp. Zool.* **282**, 677–690 (1998).
3. Martin, G. R. The roles of FGFs in the early development of vertebrate limbs. *Genes Dev.* **12**, 1571–1586 (1998).
4. Bei, M. & Maas, R. FGFs and BMP4 induce both *Msx1*-independent and *Msx1*-dependent signaling pathways in early tooth development. *Development* **125**, 4325–4333 (1998).
5. Shamim, H. *et al.* Sequential roles for Fgf4, En1 and Fgf8 in specification and regionalisation of the midbrain. *Development* **126**, 945–959 (1999).
6. Dodd, J., Jessell, T. M. & Placzek, M. The when and where of floor plate induction. *Science* **282**, 1654–1657 (1998).
7. Lee, K. J. & Jessell, T. M. The specification of dorsal cell fates in the vertebrate central nervous system. *Annu. Rev. Neurosci.* **22**, 261–294 (1999).
8. Liem, K. F. Jr, Tremml, G., Roelink, H. & Jessell, T. M. Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. *Cell* **82**, 969–979 (1995).
9. Liem, K. F. Jr, Tremml, G. & Jessell, T. M. A role for the roof plate and its resident TGF β -related proteins in neuronal patterning in the dorsal spinal cord. *Cell* **91**, 127–138 (1997).
10. Lee, K. J., Mendelsohn, M. & Jessell, T. M. Neuronal patterning by BMPs: a requirement for GDF7 in the generation of a discrete class of commissural interneurons in the spinal cord. *Genes Dev.* **12**, 3394–3407 (1998).
11. Palmiter, R. D. *et al.* Cell lineage ablation in transgenic mice by cell-specific expression of a toxin gene. *Cell* **50**, 435–443 (1987).
12. Breitman, M. L. *et al.* Genetic ablation: targeted expression of a toxin gene causes microphthalmia in transgenic mice. *Science* **238**, 1563–1565 (1987).
13. Grieshammer, U., Lewandoski, M., Prevette, D., Oppenheim, R. W. & Martin, G. R. Muscle-specific cell ablation conditional upon Cre-mediated DNA recombination in transgenic mice leads to massive spinal and cranial motoneuron loss. *Dev. Biol.* **197**, 234–247 (1998).
14. Maxwell, I. H., Maxwell, F. & Glode, L. M. Regulated expression of a diphtheria toxin A-chain gene transfected into human cells: possible strategy for inducing cancer cell suicide. *Cancer Res.* **46**, 4660–4664 (1986).
15. Rajewsky, K. *et al.* Conditional gene targeting. *J. Clin. Invest.* **98**, 600–603 (1996).
16. Sauer, B. Inducible gene targeting in mice using the Cre/lox system. *Methods* **14**, 381–392 (1998).
17. Dietrich, P., Dragatsis, I., Xuan, S., Zeitlin, S. & Efstratiadis, A. Conditional mutagenesis in mice with heat shock promoter-driven *cre* transgenes. *Mamm. Genome* (in the press).
18. Christians, E., Campion, E., Thompson, E. M. & Renard, J. P. Expression of the HSP 70.1 gene, a landmark of early zygotic activity in the mouse embryo, is restricted to the first burst of transcription. *Development* **121**, 113–122 (1995).
19. Lyons, K. M., Hogan, B. L. & Robertson, E. J. Colocalization of BMP 7 and BMP 2 RNAs suggests that these factors cooperatively mediate tissue interactions during murine development. *Mech. Dev.* **50**, 71–83 (1995).
20. Dudley, A. T. & Robertson, E. J. Overlapping expression domains of bone morphogenetic protein family members potentially account for limited tissue defects in BMP7 deficient embryos. *Dev. Dyn.* **208**, 349–362 (1997).
21. Parr, B. A., Shea, M. J. & McMahon, A. P. Mouse Wnt genes exhibit discrete domains of expression in its early embryonic CNS and limb buds. *Development* **119**, 247–261 (1993).

22. Hill, R. E. *et al.* A new family of mouse homeo box-containing genes: molecular structure, chromosomal location, and developmental expression of Hox-7.1. *Genes Dev.* **3**, 26–37 (1989).
23. Robert, B., Sassoon, D., Jacq, B., Gehring, W. & Buckingham, M. Hox7, a mouse homeobox gene with a novel pattern of expression during embryogenesis. *EMBO J.* **8**, 91–100 (1989).
24. Nagai, T. *et al.* The expression of the mouse *Zic1*, *Zic2*, and *Zic3* genes suggests an essential role for *Zic* genes in body pattern formation. *Dev. Biol.* **182**, 299–313 (1997).
25. Dickinson, M. E., Selleck, M. A., McMahon, A. P. & Bronner, F. M. Dorsalization of the neural tube by the non-neural ectoderm. *Development* **121**, 2099–2106 (1995).
26. Selleck, M. & Bronner-Fraser, M. Origins of the avian neural crest; the role of neural plate–epidermal interactions. *Development* **121**, 525–538 (1995).
27. Labosky, P. A. & Kaestner, K. H. The winged helix transcription factor *Hfh2* is expressed in neural crest and spinal cord during mouse development. *Mech. Dev.* **76**, 185–190 (1998).
28. Ma, Q., Fode, C., Guillemot, F. & Anderson, D. J. Neurogenin1 and neurogenin2 control two distinct waves of neurogenesis in developing dorsal root ganglia. *Genes Dev.* **13**, 1717–1728 (1999).
29. Jones, C. M., Lyons, K. M. & Hogan, B. L. Involvement of Bone Morphogenetic Protein-4 (BMP-4) and *Vgr-1* in morphogenesis and neurogenesis in the mouse. *Development* **111**, 531–542 (1991).
30. Arkell, R. & Beddington, R. S. P. BMP7 influences pattern and growth of the developing hindbrain of mouse embryos. *Development* **124**, 1–12 (1997).
31. Roelink, H. & Nusse, R. Expression of two members of the Wnt family during mouse development: restricted temporal and spatial patterns in the developing neural tube. *Genes Dev.* **5**, 381–388 (1991).
32. Snow, D. M., Steindler, D. A. & Silver, J. Molecular and cellular characterization of the glial roof plate of the spinal cord and optic tectum: a possible role for a proteoglycan in the development of an axon barrier. *Dev. Biol.* **138**, 359–376 (1990).
33. Ericson, J., Morton, S., Kawakami, A., Roelink, H. & Jessell, T. M. Two critical periods of Sonic Hedgehog signaling required for the specification of motor neuron identity. *Cell* **87**, 661–673 (1996).
34. Dickinson, M. E., Krumlauf, R. & McMahon, A. P. Evidence for a mitogenic effect of Wnt-1 in the developing mammalian central nervous system. *Development* **120**, 1453–1471 (1994).
35. Ikeya, M., Lee, S. M. K., Johnson, J. E., McMahon, A. P. & Takada, S. Wnt signalling required for expansion of neural crest and CNS progenitors. *Nature* **389**, 966–970 (1997).
36. Ericson, J. *et al.* Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. *Cell* **90**, 169–180 (1997).
37. Akazawa, C., Ishibashi, M., Shimizu, C., Nakanishi, S. & Kageyama, R. A mammalian helix–loop–helix factor structurally related to the product of the *Drosophila* proneural gene *atonal* is a positive transcriptional regulator expressed in the developing nervous system. *J. Biol. Chem.* **270**, 8730–8738 (1995).
38. Helms, A. W. & Johnson, J. E. Progenitors of dorsal commissural interneurons are defined by Math1 expression. *Development* **125**, 919–928 (1998).
39. Ma, Q., Kintner, C. & Anderson, D. J. Identification of *neurogenin*, a vertebrate neuronal determination gene. *Cell* **87**, 43–52 (1996).
40. Lo, L. C., Johnson, J. E., Wuenschell, C. W., Saito, T. & Anderson, D. J. Mammalian achaete-scute homolog 1 is transiently expressed by spatially restricted subsets of early neuroepithelial and neural crest cells. *Genes Dev.* **5**, 1524–1537 (1991).
41. Bronner-Fraser, M. & Fraser, S. E. Cell lineage analysis reveals multipotency of some avian neural crest cells. *Nature* **335**, 161–164 (1988).
42. O'Gorman, S., Fox, D. T. & Wahl, G. M. Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251**, 1351–1355 (1991).
43. Tsien, J. Z. *et al.* Subregion- and cell type-restricted gene knockout in mouse brain. *Cell* **87**, 1317–1326 (1996).
44. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
45. Yamaizumi, M., Mekada, E., Uchida, T. & Okada, Y. One molecule of diphtheria toxin fragment A introduced into a cell can kill the cell. *Cell* **15**, 245–250 (1978).
46. Millonig, J. H., Millen, K. J. & Hatten, M. E. The mouse *dreher* gene (*Lmx1a*) controls formation of the roof plate in the vertebrate CNS. *Nature* **403**, 764–769 (2000).
47. Tanabe, Y. & Jessell, T. M. Diversity and pattern in the developing spinal cord. *Science* **274**, 1115–1123 (1996).
48. Ding, Q. *et al.* Diminished Sonic hedgehog signaling and lack of floor plate differentiation in *Gli2* mutant mice. *Development* **125**, 2533–2543 (1998).
49. Matisse, M. P., Epstein, D. J., Park, H. L., Platt, K. A. & Joyner, A. L. *Gli2* is required for induction of floor plate and adjacent cells, but not most ventral neurons in the mouse central nervous system. *Development* **125**, 2759–2770 (1998).
50. Beattie, C. E. *et al.* Temporal separation in the specification of primary and secondary motoneurons in zebrafish. *Dev. Biol.* **187**, 171–182 (1997).

Acknowledgements

We thank A. Efstratiadis for the *Hs-cre* strain; B. Han and M. Mendelsohn for help with cell culture and generation of mouse strains; S. Kaplan for technical assistance; D. Anderson, J. Johnson, K. Kaestner, A. Pierani, P. Soriano and E. Turner for mouse strains and reagents; J. Millonig, K. Millen and M. E. Hatten for discussions; R. Axel, J. Briscoe, N. Shah and L. Voss for comments on the manuscript; and K. MacArthur for help in its preparation. K.J.L. was an HHMI Fellow of the Life Sciences Research Foundation. The work of P.D. was supported by an NIH grant to A. Efstratiadis. T.M.J. was supported by grants from the NIH and is an Investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to T. M. J. (e-mail: tmj1@columbia.edu).

Origin of the Moon's orbital inclination from resonant disk interactions

William R. Ward & Robin M. Canup

Southwest Research Institute, 1050 Walnut Street, Suite 426, Boulder, Colorado 80302, USA

The Moon is generally believed to have formed from the debris disk created by a large body colliding with the early Earth^{1,2}. Recent models of this process predict that the orbit of the newly formed Moon should be in, or very near, the Earth's equatorial plane^{3,4}. This prediction, however, is at odds with the known history of the lunar orbit: the orbit is currently expanding, but can be traced back in time to reveal that, when the Moon formed, its orbital inclination relative to the Earth's equator was $I \approx 10^\circ$ (refs 5, 6). The cause of this initial inclination has been a mystery for over 30 years, as most dynamical processes (such as those that act to flatten Saturn's rings) will tend to decrease orbital inclinations. Here we show that the Moon's substantial orbital inclination is probably a natural result of its formation from an impact-generated disk. The mechanism involves a gravitational resonance between the Moon and accretion-disk material, which can increase orbital inclinations up to $\sim 15^\circ$.

The Moon-forming impact is believed to have produced a disk of debris^{7–9} that was roughly centred at the Earth's Roche limit, $a_{\text{Roche}} \sim 2.9R_\oplus$ and $R_\oplus$ is the radius of the Earth (6,378 km). The Roche limit is the distance interior to which a fluid body with no internal strength will be disrupted by a planet's gravity. Inside a_{Roche} , collisional aggregation is inhibited by the Earth's tidal forces and orbiting material remains dispersed in a disk; consequently, it is the portions of the disk exterior to a_{Roche} that rapidly accumulate to become the Moon^{3,4}. Accretion studies⁴ find that producing an object the size of the Moon with an orbital radius of $a \approx 1.2a_{\text{Roche}} \approx 3.5R_\oplus$ requires an initial disk containing at least two lunar masses, or $M_{\text{disk}} \approx 2M_{\text{Moon}}$. Such studies typically yield a moon with $I \approx 1^\circ$. To date, only two explanations for the cause of the initial $\sim 10^\circ$ lunar inclination appear viable, and these require either another large impact improbably soon after the Moon-forming event, or multiple passes of the Moon through two orbital resonances involving the Sun under a rather specialized set of conditions¹⁰. The resonant interaction we identify here would precede either of these potential events, and is effective for a wide range of plausible conditions.

Estimates of the lifetime of an impact-generated protolunar disk have varied. Recent simulations⁴ are computationally limited to describing the disk as several thousand "moonlets", each initially hundreds of kilometres in size. Such a disk persists for only about a year until objects are accreted by the growing Moon or scattered onto the Earth. However, moonlet-sized objects could not form within a_{Roche} to begin with, and so this short time estimate is probably applicable only to that portion of the disk exterior to the Roche limit. An inner disk composed of small debris would be more collisionally dominated and resistant to rapid depletion via scattering by the Moon than accretion simulations⁴ suggest. Such a disk would respond to gravitational perturbations from a satellite in a collective fashion (by, for example, the generation of spiral waves), rather than through discrete scattering events. In addition, the rate of disk spreading (proportional to the disk's viscosity) predicted by the purely dynamical estimates^{11,12} is so large that the corresponding energy release rate greatly exceeds the disk's radiative cooling rate¹³. A more appropriate limit would be a viscosity regulated by the disk cooling time, $\sim 10^2$ years.

These results suggest that the Moon would have co-existed for

some time with an inner disk containing a good fraction of a lunar mass. The inclination-generating mechanism we propose involves a resonance between this inner remnant disk and the Moon. Disk-satellite resonances are well-studied processes associated with, for example, spiral waves observed in Saturn's rings¹⁴ and the shepherding of the uranian ringlets¹⁵.

Periodic perturbations by a satellite excite waves in a companion disk at locations where the ratio of the orbital period in the disk to that of the satellite is approximately equal to a ratio of two integers. Mean motion resonances that involve in-plane perturbations are known as Lindblad resonances; vertical resonances arise from perturbations perpendicular to the plane of the disk. The strongest resonances are zero-order, independent of either e (eccentricity) or I (inclination); the next strongest are first-order resonances which have a linear dependence on e or I . In an interior disk, the first-order $m = 2$ inner Lindblad resonance (ILR) occurs at the location in the disk where the orbital period is about one-third that of the satellite;

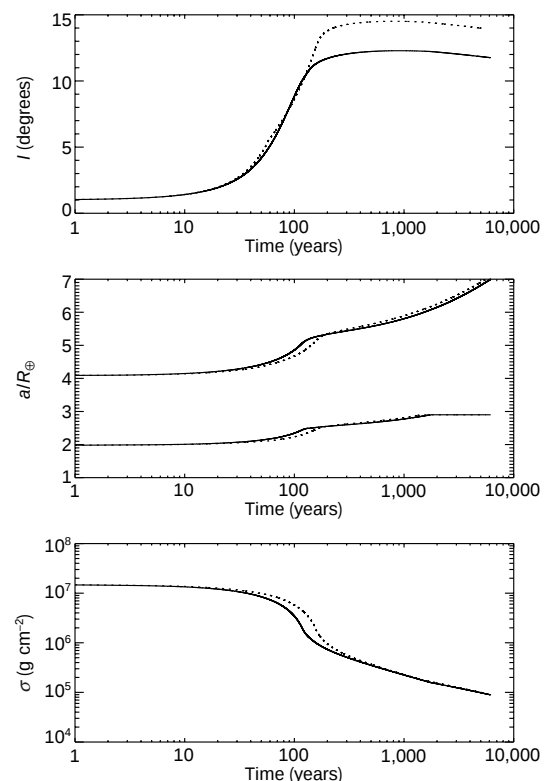


Figure 1 The evolution of the Moon's orbit as it resonantly interacts with an inner impact-generated disk. Values of inclination (I ; top panel), radius ($a/R_\oplus$; middle panel), and disk surface density (σ ; bottom panel) are shown as the Moon recoils away from a disk with $M_{\text{disk}}(0) = 0.75M_{\text{Moon}}$. Data are shown for $t_{\text{spread}}(0) = 37$ years (solid lines) and 50 years (dotted lines). We assume that the Moon forms outside the Roche distance, and for some time co-exists with a disk that initially extended from $R_\oplus$ to $r_{\text{disk}}(0) = a_{\text{Roche}}$. The initial angular momentum of the system is defined by $M_{\text{disk}}(0)$, and the assumption that material which comprises the Moon had the same starting angular momentum as a lunar mass orbiting at $\sim 1.2a_{\text{Roche}} \sim 3.5R_\oplus$ (based on results to date of accretion simulations^{3,4}, although I_{max} is not overly sensitive to this choice). The disk edge (lower curves in middle panel) contracts to $r_{\text{disk}} \approx 2R_\oplus$ by the time the Moon's orbit (upper curves in middle panel) expands to $\sim 4.1R_\oplus$ (due to zero-order and first order ILR torques) and the $m = 2$ ILR becomes the only first-order mean-motion resonance left in the disk. (We note that higher-order resonances also in the disk at this time will be rendered ineffectual by rapid eccentricity-damping by $m = 1$ secular ILR^{20,21}.) From $4.1R_\oplus$, the Moon's orbit is evolved due to the $m = 2$ ILR (if it can still be felt by the disk), the $m = 2$ IVR, and terrestrial and lunar tides (we use dissipation functions $Q_\oplus = Q_{\text{Moon}} = 50$, and tidal Love numbers $k_{2\oplus} = 10k_{2\text{Moon}} = 0.25$). The discontinuity at ~ 120 years is due to the switch from a radiation-limited viscosity¹³ (equation (5)) to an instability-driven viscosity¹¹ (equation (6)).

here m corresponds to the number of spiral arms in the resulting density wave pattern. For an inclined satellite, spiral bending waves (which corrugate the disk) are launched at inner vertical resonances (IVRs)¹⁴. For resonances with $m > 1$, the interaction of a satellite with waves it generates in an interior disk causes a transfer of energy and angular momentum from the disk (which contracts) to the satellite (whose orbit expands). For a non-circular and/or inclined satellite orbit, disk torques also alter e and/or I (refs 16–18). The torque values applicable for a given resonance have been shown to be quite insensitive to the particular physical state of the disk material, such as its viscosity, pressure and self-gravity¹⁹.

As a moon (or moons) first accumulates outside a_{Roche} , there would be many resonances within a disk extending from the Earth's surface to the Roche limit. Torques from the strongest resonances would cause satellite-sized bodies to migrate outward at rates comparable to their month-to-year accumulation time⁴, t_{acc} . Satellite migration would be accompanied by a migration of the mean motion resonances (which each occur at a fixed fraction of a satellite's position, a); these resonances would pass outward through the disk, and finally leave it. As the strongest, zero-order resonances migrate out of the disk, the expansion rate of the moon(s) slows. In particular, when the orbital radius reaches $a \approx 3^{2/3} r_{\text{disk}} \approx 2.08 r_{\text{disk}}$, where r_{disk} is the radius of the disk's outer edge, the first-order $m = 2$ ILR leaves the disk, and the migration time for a lunar-sized body slows to a value that is much longer than t_{acc} . At this point, the $m = 2$ IVR—which is located slightly interior¹⁴ to the $m = 2$ ILR because of dissimilar precession rates of the orbital apse and node—becomes the only first-order mean-motion resonance remaining in the disk, and acts to increase the lunar inclination. Any eccentricity generated during the preceding recoil phase is quickly damped by the $m = 1$ secular resonance which resides in the disk interior to the $m = 2$ IVR (refs 20, 21).

The $m = 2$ IVR increases a and I , at the rates^{14,16,17}:

$$\frac{da}{dt} = \frac{3T_{\text{IVR}}}{Ma\Omega} \quad \text{and} \quad \frac{dI}{dt} = \frac{T_{\text{IVR}}}{Ma^2\Omega \sin I} \left(\frac{3}{2} \cos I - 1 \right) \quad (1)$$

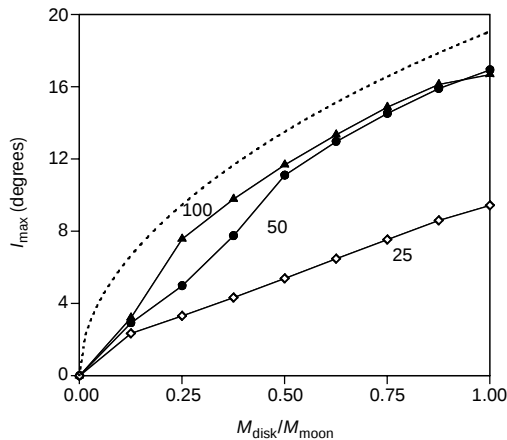


Figure 2 The Moon's initial orbital inclination. Maximum lunar orbital inclinations, I_{max} , as a function of $M_{\text{disk}}(0)$, are given for disk spreading times (t_{spread}) of 25, 50 and 100 years and assuming that $a_0 = 3.5R_{\oplus}$. A lower-mass disk yields a lower I_{max} for a given value of t_{spread} due to the dependence of the IVR torque on disk surface density. For a given M_{disk} , shortening the disk spreading time also decreases I_{max} . In this case, the time for r_{disk} to reach a_{Roche} or for the disk to be depleted (and thus the time over which the Moon experiences the IVR torque) is reduced. The dashed line is the analytical upper limit for I_{max} , which neglects solid body tides and assumes that the IVR remains in the disk until the disk loses all of its mass to the Earth. In this case, the final angular momentum of the Moon equals the initial angular momentum in the disk and the Moon, minus that absorbed by the Earth as it accretes the disk, that is, $L_{\text{absorb}} = M_{\text{disk}}(GM_{\oplus}R_{\oplus})^{1/2}$.

where M is the mass of the moon, Ω is its orbital frequency, and T_{IVR} is the torque on the moon due to the $m = 2$ IVR. While the IVR is in the disk, T_{IVR} is given by

$$T_{\text{IVR}} = \sin^2 I \frac{2\pi^2 \sigma}{3\Omega^2} \left(\frac{GM}{2a} \right)^2 [\alpha_{\text{IVR}}^{5/2} b_{3/2}^2(\alpha_{\text{IVR}})]^2 \quad (2)$$

where σ is the disk surface mass density, G is the universal gravitational constant, a_{IVR} is the orbital radius of the IVR, α_{IVR} is the ratio $(a_{\text{IVR}}/a) \approx 3^{-2/3}$, and $b_{3/2}^2(\alpha_{\text{IVR}})$ is a Laplace coefficient²². Assuming that the $m = 2$ IVR is the sole source of lunar migration and that $\cos I_0 \approx 1$, the rates in equation (1) can be divided and integrated to find

$$\cos I_{\text{max}} = \frac{2}{3} + \frac{1}{3} \sqrt{\frac{a_1}{a_2}} \quad (3)$$

where I_{max} is the maximum inclination produced by the IVR, and a_1 and a_2 are the lunar orbital radii when the first-order $m = 2$ ILR and the $m = 2$ IVR leave the disk, respectively. An estimate of a_1 is found by requiring conservation of mass, and that the sum of the angular momenta of the disk and the moon is equal to the starting angular momentum of the system. This is valid (independent of the details of earlier moon(s)–disk interactions) if the initial recoil is much faster than the disk can viscously spread, or than the moon can respond to solid body tides. The largest value of a_2 (and the largest lunar inclination) is achieved if the IVR remains in the disk while it spreads until the disk has lost all of its mass to the Earth.

A simple simulation illustrates the Moon's orbital evolution as it recoils from its precursor disk. We do not model the accretion phase, but assume that the Moon assembled quickly at an orbital radius a_0 and has migrated outward while the disk contracts until $a(t) = 2.08 r_{\text{disk}}(t)$. Recent impact simulations⁹ yield debris disks with total masses $\sim 2M_{\text{Moon}}$ and total angular momenta $\sim 0.3J_{\oplus-M}$, where $J_{\oplus-M}$ is the current angular momentum of the Earth–Moon system. Such a disk yields a Moon on a circular orbit with $a_0 = 3.4R_{\oplus}$ (ref. 4); somewhat larger a_0 values are possible given non-zero eccentricities for the Moon and inner disk debris. Accretion simulations⁴ typically find ($a_{\text{Roche}} \leq a_0 \leq 1.5a_{\text{Roche}}$); for all values in this range, the $m = 2$ IVR is within the disk when the Moon forms.

We track the evolution of the lunar orbit as it evolves due to (1) the $m = 2$ ILR and IVR, (2) interaction with tides it raises on the Earth^{5,6}, and (3) energy dissipation due to lunar tides raised by the Earth¹⁰. The disk evolves due to both loss of angular momentum to the Moon via the $m = 2$ ILR and IVR, and viscous diffusive spreading. For the latter, we use a time-dependent disk viscosity, $\nu(t)$, equivalent to that predicted by dynamical simulations¹² and analytical estimates¹¹, unless this value implies an energy release rate that exceeds the disk's cooling rate. In the latter case, we assume a balance between these rates, so that $\sigma_{\text{S-B}} T^4 = \frac{9}{8} \nu \sigma \Omega_{\text{disk}}^2$, where $\sigma_{\text{S-B}}$ is the Stefan–Boltzmann constant, T is the disk photospheric temperature, and Ω_{disk} is the orbital frequency at r_{disk} (ref. 13). This balance yields a radiation-limited minimum disk spreading time in years, t_{spread} , given by (ref. 13):

$$t_{\text{spread}} = \frac{r_{\text{disk}}^2}{\nu} \approx 50 \left(\frac{r_{\text{disk}}}{a_{\text{Roche}}} \right)^{-3} \left(\frac{T}{2,000 \text{ K}} \right)^{-4} \left(\frac{M_{\text{disk}}}{M_{\text{Moon}}} \right) \quad (4)$$

The initial values of $t_{\text{spread}}(0)$ and $M_{\text{disk}}(0)$ (that is, those when $r_{\text{disk}} = a_{\text{Roche}}$) are input parameters. At each time step, $\nu(t)$ is updated to be the smaller of two values:

$$\nu(t) = \nu(0) \left(\frac{M_{\text{disk}}(0)}{M_{\text{disk}}(t)} \right) \left(\frac{r_{\text{disk}}(t)}{a_{\text{Roche}}} \right)^5 \quad (5)$$

which is the radiation-limit for $\nu(t)$ from equation (4), or

$$\nu(t) = \left(\frac{M_{\text{disk}}(t)}{M_{\oplus}} \right)^2 \Omega_{\text{disk}} r_{\text{disk}}^2(t) \quad (6)$$

which is the viscosity obtained by a disk stability estimate¹¹. The disk mass is decreased as material spreads onto the Earth or past a_{Roche} .

Figure 1 shows the evolution of I , a and σ as the Moon recoils from a disk of mass $M_{\text{disk}} = 0.75 M_{\text{Moon}}$, assuming that $a_0 = 3.5R_{\oplus}$ and $a_1 = 4.1R_{\oplus}$. The resulting maximum inclinations are 12.3° and 14.5° for $t_{\text{spread}} = 37$ and 50 years, respectively. The timescale, t_{IVR} , for growth in a due to the IVR is regulated by the location of the disk edge to be comparable to t_{spread} . As the IVR nears the edge of the disk, the back torque from the Moon retards the disk's tendency to spread, and the IVR begins to migrate out of the disk. But as this occurs, the back torque on the disk is decreased, allowing increased spreading of disk material outward, until once again the IVR is in the disk. In this manner, the location of $r_{\text{disk}}(t)$ and the location of the IVR move in lock step. Once $r_{\text{disk}} = a_{\text{Roche}}$, the edge of the continuous disk does not advance because material spreading beyond a_{Roche} can accrete into discrete objects which are not effective at sustaining wave action. The continued outward migration of the Moon then causes the IVR to leave the disk, preventing further resonant growth of I ; the subsequent system evolution is determined by tidal interaction with the Earth and Sun^{5,6,10}. Figure 2 shows I_{max} resulting from a variety of choices of $M_{\text{disk}}(0)$ and $t_{\text{spread}}(0)$.

Until now, an apparent deficiency of the impact hypothesis has been its inability to account for the inclination of the lunar orbit²³. Here we identify a resonance that would naturally result as the Moon recoiled from an interior disk; this resonance could, for a range of plausible conditions, yield the requisite lunar inclination. The resulting inclination depends mainly on two parameters: $M_{\text{disk}}(0)$ and $t_{\text{spread}}(0)$. Simulations of potential Moon-forming impacts by Cameron⁹ yield $M_{\text{disk}} \approx 1.5\text{--}2 M_{\text{Moon}}$, with somewhat more than half of the mass exterior to a_{Roche} . Forming the Moon at the outer edge of such a disk would leave an inner remnant disk of $\sim 0.5\text{--}1 M_{\text{Moon}}$. An inner disk whose spreading rate is regulated by its radiative cooling time with a nominal photospheric temperature of 2,000 K will have $t_{\text{spread}}(0) \approx 50(M_{\text{disk}}/M_{\text{Moon}})$ years. These values coincide well with those needed to yield $I_{\text{max}} \approx 10^\circ$. The inclination of the Moon's orbit may thus represent an important clue to the early state of the post-impact system; we expect that this will motivate further modelling of the physics of an impact-generated protolunar disk. □

Received 8 September 1999; accepted 4 January 2000.

1. Cameron, A. G. W. & Ward, W. R. The origin of the Moon. *Lunar Sci.* VII, 120–122 (1976).
2. Hartmann, W. K. & Davis, D. R. Satellite-sized planetesimals and lunar origin. *Icarus* **24**, 504–515 (1975).
3. Canup, R. M. & Esposito, L. W. Accretion of the Moon from an impact-generated disk. *Icarus* **119**, 427–446 (1996).
4. Ida, S., Canup, R. M. & Stewart, G. Formation of the Moon from an impact-generated disk. *Nature* **389**, 353–357 (1997).
5. Goldreich, P. History of the lunar orbit. *Rev. Geophys.* **4**, 411–439 (1966).
6. Touma, J. & Wisdom, J. Evolution of the Earth–Moon system. *Astron. J.* **108**, 1943–1961 (1994).
7. Cameron, A. G. W. & Benz, W. The origin of the Moon and the single impact hypothesis IV. *Icarus* **92**, 204–216 (1991).
8. Cameron, A. G. W. The origin of the Moon and the single impact hypothesis V. *Icarus* **126**, 126–137 (1997).
9. Cameron, A. G. W. & Canup, R. M. The giant impact occurred during Earth accretion. *Lunar Planet. Sci.* **XXIX**, 1062–1063 (1998).
10. Touma, J. & Wisdom, J. Resonances in the early evolution of the Earth–Moon system. *Astron. J.* **115**, 1653–1663 (1998).
11. Ward, W. R. & Cameron, A. G. W. Disc evolution within the Roche limit. *Lunar Planet. Sci.* **IX**, 1205–1207 (1978).
12. Kokubo, E., Ida, S. & Makino, J. Evolution of a circumterrestrial disk and formation of a single moon. *Icarus* (submitted).
13. Thompson, C. & Stevenson, D. Gravitational instability in two-phase disks and the origin of the Moon. *Astrophys. J.* **333**, 542–481 (1983).
14. Shu, F. in *Planetary Rings* (eds Greenberg, R. & Brahic, A.) 513–561 (Univ. Arizona Press, Tucson, 1984).
15. Elliot, J. L. & Nicholson, P. D. in *Planetary Rings* (eds Greenberg, R. & Brahic, A.) 25–72 (Univ. Arizona Press, Tucson, 1984).
16. Goldreich, P. & Tremaine, S. Disk-satellite interactions. *Astrophys. J.* **241**, 425–441 (1980).
17. Borderies, N., Goldreich, P. & Tremaine, S. Excitation of inclinations in ring-satellite systems. *Astrophys. J.* **284**, 429–434 (1984).
18. Ward, W. R. & Hahn, J. M. Damping of orbital inclinations by bending waves. *Icarus* **110**, 95–108 (1994).

19. Meyer-Vernet, N. & Sicardy, B. On the physics of resonant disk-satellite interaction. *Icarus* **69**, 157–175 (1987).
20. Ward, W. R. & Hahn, J. M. The eccentricity of Neptune and the nature of the Kuiper belt. *Science* **280**, 2104–2106 (1998).
21. Tremaine, S. Resonant relaxation in protoplanetary disks. *Astron. J.* **116**, 2015–2022 (1998).
22. Brouwer, D. & Clemence, G. M. *Methods of Celestial Mechanics* (Academic, New York, 1961).
23. Boss, A. P. & Peale, S. J. in *Origin of the Moon* (eds Hartmann, W. K., Phillips, R. J. & Taylor, G. J.) (Lunar and Planetary Institute Press, Houston, 1986).

Acknowledgements

We thank L. Dones, D. Stevenson, G. Stewart and J. Wisdom for comments, and A. G. W. Cameron for data from his SPH impact simulations. This work was supported by NASA's Planetary Geology and Geophysics, and Origins of Solar Systems, Programs.

Correspondence and requests for materials should be addressed to W.R.W. (e-mail ward@boulder.swri.edu).

Preparing pure photon number states of the radiation field

B. T. H. Varcoe*, S. Brattke*†, M. Weidinger* & H. Walther*†

* Max Planck Institute for Quantum Optics, Hans-Kopfermann Strasse 1, D-85748 Garching, Germany

† Sektion Physik der Universität München, 85748 Garching, Germany

The quantum mechanical description of a radiation field is based on states that are characterized by the number of photons in a particular mode; the most basic quantum states are those with fixed photon number, usually referred to as number (or Fock) states. Although Fock states of vibrational motion can be observed readily in ion traps¹, number states of the radiation field are very fragile and difficult to produce and maintain. Single photons in multi-mode fields have been generated using the technique of photon pairs^{2,3}. But in order to generate these states in a cavity, the mode in question must have minimal losses; moreover, additional sources of photon number fluctuations, such as the thermal field, must be eliminated. Here we observe the build-up of number states in a high-Q cavity, by investigating the interaction dynamics of a probe atom with the field. We employ a dynamical method of number state preparation that involves state reduction of highly excited atoms in a cavity, with a photon lifetime as high as 0.2 seconds. (This set-up is usually known as the one-atom maser or 'micromaser'.) Pure states containing up to two photons are measured unambiguously.

Number states are also generated as trapping states in a micromaser⁴; however, it is not easy to study the purity of the generated states in that case. There are many experiments where single-photon exchange between subsequent atoms has been investigated (see, for example, refs 5 and 6) leading to an entanglement between subsequent atoms and possibly the photons. In those experiments, no state reduction of the photon field in the cavity is performed, and so the field itself is in a more complicated state. During the review process of this Letter, work⁷ describing quantum non-demolition measurement of states containing an average of one or zero photons in the mode related to a parity measurement of the photon state was published (see ref. 8 for comparison).

In the micromaser, highly excited atoms (usually called Rydberg atoms) interact with a single mode of a superconducting cavity with a quality factor as high as 3×10^{10} , leading to the above-mentioned photon lifetime. The steady-state field generated in the micromaser has been the subject of many detailed studies, such as the observation of sub-poissonian photon statistics⁹. The dynamics of the atom-field photon exchange has also been investigated in the

collapse and revivals of the Rabi nutation¹⁰, atom interference¹¹, bistability and quantum jumps of the field¹², atom–field and atom–atom entanglement¹³.

The generation of number states in the micromaser has long been anticipated^{13,14}. In this connection trapping states were of interest, as, under certain conditions, the ‘trapped’ steady-state field represents number states. We briefly discuss here the essential features of trapping states of the micromaser field, as our results will also be compared with those expected for the trapping situation.

The interaction of a two-level atom with a single mode of the cavity field is the model system described by the Jaynes–Cummings hamiltonian¹⁵. In this model, an atom in the presence of a resonant quantum field undergoes Rabi oscillations. Trapping states occur when the atom–field coupling, Ω , and the interaction time, t_{int} , are chosen such that in a cavity field with n photons each atom undergoes an integer number, k , of Rabi cycles. This is summarized by the condition:

$$-t_{\text{int}}\sqrt{n+1} = k\pi \quad (1)$$

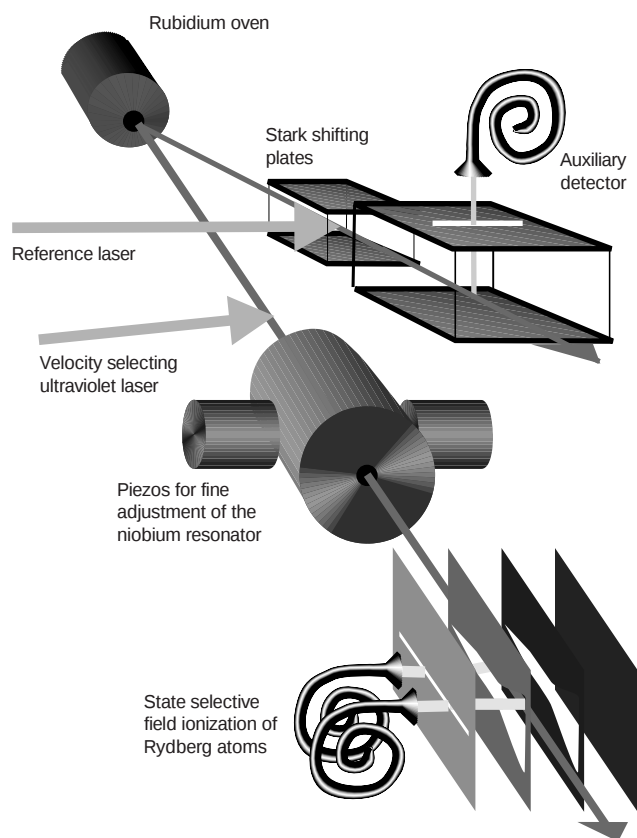


Figure 1 The experimental set-up. A frequency-doubled dye laser ($\lambda = 297$ nm) was used to excite rubidium ^{85}Rb atoms to the Rydberg $63\text{P}_{3/2}$ state from the $5\text{S}_{1/2}$ ($F = 3$) ground state. The maser cavity is tuned to the $63\text{P}_{3/2}$ – $61\text{D}_{5/2}$ (21.456 GHz) transition and tuning of the cavity is performed with two piezo translators. Velocity selection is provided by angling the excitation laser towards the main atomic beam at 11° to the normal. The dye laser was locked to a reference beam, using an external computer control, to the $5\text{S}_{1/2}$ – $63\text{P}_{3/2}$ transition of the reference atomic beam excited under normal incidence. In order to obtain controlled continuous tuning of the laser frequency, and hence the atomic velocity, a Stark field was applied to the reference beam. This Stark field was produced by means of a high-quality programmable power supply. The cavity is cooled below 300 mK by the cryostat, corresponding to a maximum thermal photon number of 0.033. For interaction times below $40 \mu\text{s}$ the excitation spectra of $63\text{P}_{1/2}$ and $63\text{P}_{3/2}$ levels overlap, leading to excitation of $63\text{P}_{1/2}$ atoms which do not interact in the cavity field; however, they are counted in the detectors leading to a perturbation of the counting statistics. The Rydberg atoms are detected by field ionization in two detectors set at different voltages so that the upper and lower states, $63\text{P}_{3/2}$ and $61\text{D}_{5/2}$, can be counted separately.

When this condition is satisfied, the photon number is left unchanged following the interaction of an atom with the field and the photon number is therefore trapped¹⁶. It has been demonstrated⁴ that trapping states up to $n = 5$ are present. In the trapping states, whenever a photon is lost from the field due to dissipation, the next excited-state atom entering the cavity will emit a photon with a high probability and restore the photon number.

Here we report on an alternative method of generating number states. This method has the advantage that the generated states can be analysed in an unambiguous way afterwards. In the usual operation of the micromaser, atoms in the upper state of the maser transition are injected into the cavity. In the absence of dissipation, atoms detected in the lower state must have emitted a single photon into the cavity field. The interaction of an atom with a field that is initially in the vacuum state $|0\rangle$, or more generally in a state $|n\rangle$, changes the field state to a superposition of the states $|n\rangle$ and $|n+1\rangle$. By measuring an atom in the ground state, the superposition is reduced to the state $|n+1\rangle$. Therefore the determination of the state of all outgoing atoms reduces the state of the cavity field to a pure number state¹⁷. Without state reduction a mixture of the $|n\rangle$ and $|n+1\rangle$ states would persist in the cavity. The presence of dissipation complicates the discussion and as a result a fixed number state is not precisely reached (see the data analysis below).

In the presence of a cavity photon number, $|n\rangle$, the relative populations of the excited and ground states of an atom will oscillate at a frequency $\sim \sqrt{n+1}$. Experimentally the atomic inversion, given by $I = P_g - P_e$, is measured. Here P_g and P_e are the probability of finding a ground-state or excited-state atom, respectively. If the

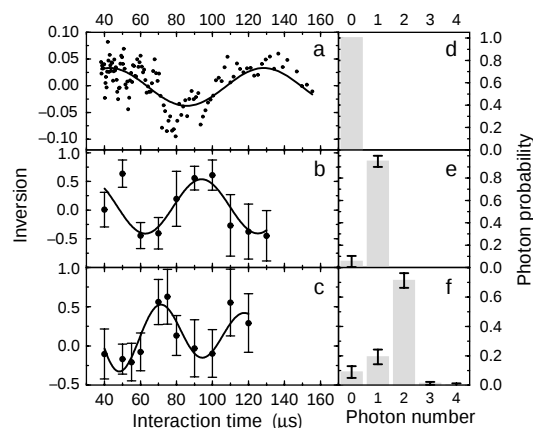


Figure 2 Three Rabi oscillations. **a**, **b** and **c** represent the number states $n=0$, 1 and 2, respectively. The plots in **d**, **e** and **f** display the coefficients P_n . The photon distribution P_n was calculated for each Rabi cycle by fitting equation (2) to each plot for the set of photon numbers $n=0$ to $n=3$. The relative phase of the Rabi frequency was fixed because all the atoms enter in the excited state of the maser transition. In each fit the highest probability was obtained for the target number state. Unlike the $n=1$ and $n=2$ Rabi cycles, the $n=0$ oscillation in **a** was obtained in the steady-state operation of the micromaser in a very low-flux regime. The fit to this curve was performed for Rabi cycles from $n=0$ to $n=2$. The low visibility of this curve was due to the low flux (<1 Hz) which was required to reduce the steady-state operation of the micromaser to below-threshold behaviour; hence, detector dark counts become comparable to the real count rates and therefore contribute to a large background. To improve the measurements for photon number of $n=3$ and higher, the range of interaction times would have to be extended beyond $120 \mu\text{s}$, but this is not possible with the current apparatus. During the Rabi cycle the cavity photon number changes periodically. At each maximum there is one photon more than at the minimum. The Rabi oscillation thus allows a non-destructive and repeated measurement of the photon number to be performed. See refs 7 and 8 for comparison. In connection with the discussion of trapping states, it is interesting to note that minima in the number-state Rabi oscillations correspond precisely to the trapping-states conditions of the steady-state field⁴. Therefore the large possible storage times of single photons would permit investigation of the transition from a pulsed to a steady-state experiment.

field is not in a number state the inversion is given by:

$$I(t_{\text{int}}) = -C \sum_n P_n \cos(2 - \sqrt{n+1} t_{\text{int}}) \quad (2)$$

Here P_n is the probability of finding n photons in the mode and t_{int} is the interaction time of the atoms with the cavity field. Rabi cycling of a probe atom can therefore be used as an unambiguous test of the photon distribution in the cavity by analysing the dynamics using equation (2). The factor C considers the reduction of the signal amplitude as a result of dark counts.

The experimental apparatus of the micromaser used for the present experiment is presented in Fig. 1. The present set-up has been described in detail^{4,12}. The natural velocity distribution of the atomic beam allowed the interaction time of the atoms with the cavity to be tuned from 40 μs to 160 μs . There was an uncertainty in the interaction time of at most 3%. The cavity used for the measurement described here had a Q factor of 3.4×10^9 , corresponding to a photon lifetime of 25 ms. A higher Q value would have increased the measuring time as discussed below.

In contrast to previous experiments with the micromaser, pulsed excitation of the atoms is used so that the number of atoms passing through the cavity can be predetermined. The probability of detecting an atom in either the excited or ground state is about 40% with a 3–5% miscount rate. Owing to the finite lifetime of the atoms, additional atoms are lost through spontaneous emission in the flight between the cavity and the detectors⁴. To create and detect an n -photon number state in the cavity, $N = n + 1$ atoms are required. That is n atoms to create the number state and the final atom as a probe of the state. However, owing to the non-perfect detector efficiency and atomic decay there are missed counts. By using a laser pulse of short duration the number of excited-state atoms entering the cavity per pulse is low. Hence we know that when N atoms per pulse were detected the probability of having $N + 1$ atoms per pulse was negligibly small. This was achieved by modifying the ultraviolet excitation pulse such that the mean number of atoms per pulse was between 0.2 and 0.8. With 40% detector efficiency and the assumption that the probability of missing a count is statistically independent, there is a probability of about 1% of the state preparation being incorrect because an atom escapes detection¹⁸. As the flux of atoms was variable, the pulse duration was also variable; a maximum sampling time of 3 ms for the $n = 1$ data and 5 ms for the $n = 2$ data was imposed to limit the time delay between the pump and probe atoms. In fact, in most cases the time delay was comparable to the excitation pulse duration. For the

measurement of an n -photon number state, the detection of the probe atom is triggered by the detection of n ground-state atoms within the length of the laser pulse. If too few or too many atoms (upper- or lower-state) are detected within the laser pulse duration, the measurement is rejected.

To ensure that the cavity is in the vacuum state at the start of a measurement, there is a delay of 1.5 cavity decay times between the laser pulses. Hence we compromised with a Q value lower than ultimately possible in our set-up, because a higher Q would lead to an increase of the data collection time. Even with the reduced cavity lifetime of 25 ms and large delay times between the laser pulses, a cyclically steady-state maser field can build up in the cavity. The time delay between pulses was selected as a compromise between limiting the growth of the maser field and the length of the data collection time.

Figure 2a–c displays three Rabi cycles obtained by measuring the inversion of a probe atom that followed the detection of $n = 0, 1$ or 2 ground-state atoms respectively. Figure 2d–f displays normalized values of the coefficients P_n obtained from these curve fits. The strongest indication that the data in Fig. 2a–c represent three number states is provided in Fig. 3, where the frequency of single sine wave curve fit is plotted for each graph alongside the theoretical variation of the Rabi frequency as a function of the photon number. There is a small dependence of the coupling constant on atom–field detuning and our day-to-day repeatability of the detuning is limited to a few kilohertz. The photon probabilities in Fig. 2d–f therefore use the frequency plotted in Fig. 3 as a reference frequency to calculate the frequencies for other possible number-state components. A similar method¹⁹, using the interaction between the atoms and field, was used to analyse a coherent microwave field injected into a cavity.

The simulations, which are described later, demonstrate that we produced number states with a purity of 99% for the $n = 1$ state and 95% for the $n = 2$ state. The fact that we do not measure pure number states is caused by dissipation in the time interval between production and analysis of the cavity field.

Because of the long waiting times for three atom events, the $n = 2$ Rabi data were more difficult to collect than the other two measurements. The data collection time became substantially longer as the interaction time was increased and background effects have a higher impact on the data. The fit to the $n = 2$ data includes an exponentially decreasing weight, so that measurements obtained for longer interaction times have less significance than those at short times.

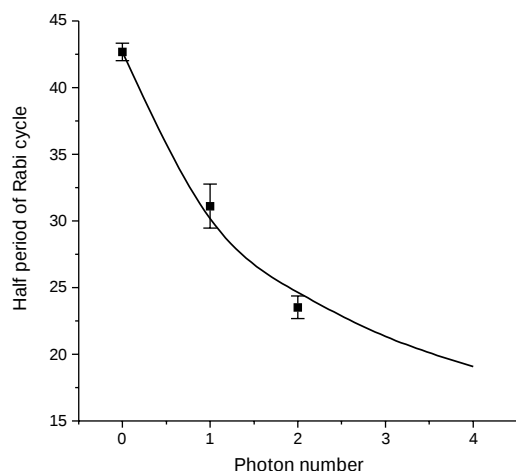


Figure 3 Dependence of the Rabi frequency on the photon number. A single sine fit to each of the three curves in Fig. 2a–c is plotted versus the theoretical variation of frequency as a function of photon number. The theoretical plot was calculated for the coupling constant of 36.8 krad s^{-1} , which was the best fit to the data.

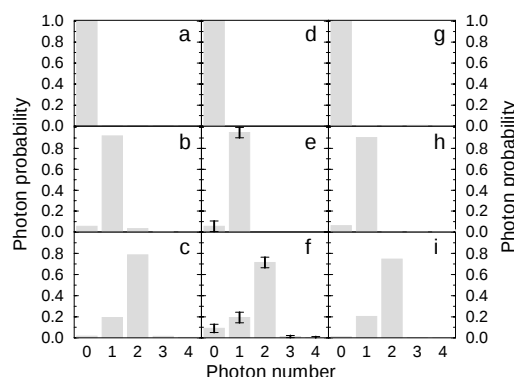


Figure 4 Theoretical and experimental results on the purity of number states. **a–c**, A theoretical simulation of the current experiment; **d–f**, the current experimental results; **g–i**, a theoretical model that extends the current experiment to the steady state at the positions of the trapping states. The agreement between the two theoretical results and the experimental result is remarkable, indicating that dissipation is the most likely loss mechanism. Without dissipation (that is, in the moment of generation) the purity of the states is 99% for $n = 1$ and 95% for $n = 2$.

The limited resolution of the $n = 1$ and $n = 2$ Rabi oscillations is a result of the cumulative effect of velocity and detuning fluctuations, a 3–5% miscount rate in the state-selective detectors and detector dark counts. The presence of number states, however, can be determined independently of the resolution because the only important factor is the accumulated Rabi phase of the probe atom as it passes through the cavity. The data for the $n = 1$ and $n = 2$ Rabi oscillations were evaluated considering possible disturbances by trapping states.

We now calculate the dissipation of the field before detection in order to determine the influence of dissipation. For these simulations we make two idealizing assumptions: thermal photons are only taken into account for the long-term build-up of the cyclically steady state, and gaussian averaging over velocity spread of atoms is considered to be about 3%. Considered in the calculations are the exponential decays for the cavity field during the pulse when either one photon (for $n = 1$) or two photons were deposited one by one (for $n = 2$) changing the photon number distribution. The simulations also average over the poissonian arrival times of the atoms. The details of this calculation will be discussed elsewhere (B. T. H. V. *et al.*, manuscript in preparation). The results of these calculations are compared to the experimental results in Fig. 4a–f.

As dissipation is the dominant loss mechanism, it is interesting to compare the purity of the number states generated by the current method with that expected for trapping states (Fig. 4g–i). The agreement of the purity of the number states is very good. The trapping-state photon distribution is generated in the steady state, which means that whenever the loss of a photon occurs the next incoming atom will restore the old field with a high probability. The non-zero amplitudes of the states $|0\rangle$ in Fig. 4h and $|1\rangle$ in Fig. 4i are due to dissipative losses before restoration of a lost photon, which is not replaced immediately but after a time interval dependent on the atom flux. The atom rate used in these calculations was 25 atoms per cavity decay time, or an average delay of 1 ms. This can be compared to the delay between the preparation and probe atoms in the present experiment. In the steady-state simulation loss due to cavity decay determines the purity of the number state; in the limit of zero loss the state measurement is perfect. We conclude that dissipative loss due to cavity decay in the delay to a probe atom largely determines the measured deviation from a pure number state. The thermal field also influences the photon distribution. The nature of the selection process in the current experiment means that we can reduce the influence of the thermal field by only performing measurements of the field state after a trigger of n ground state atoms. Hence, the state of the field is well known. The steady-state simulation was therefore performed for a temperature of 100 mK, which makes the influence of the thermal field in the steady-state correspondingly low.

The fact that number states can be readily created presents many opportunities. For example, in connection with the investigation of quantum information, the decoherence of ‘Schrödinger cat’ states can be largely attributed to dissipation. The very low influence of dissipation in this experiment provides an excellent environment for investigating decoherence and non-local quantum phenomena such as entanglement of atoms.

M

Received 26 April; accepted 2 December 1999.

- Leibfried, D. *et al.* Experimental determination of the motional quantum state of a trapped atom. *Phys. Rev. Lett.* **77**, 4281–4285 (1996).
- Hong, C. K. & Mandel, L. Experimental realization of a localized one photon state. *Phys. Rev. Lett.* **56**, 58–60 (1986).
- Fearn, H. & Loudon, R. Theory of two-photon interference. *J. Opt. Soc. Am. B* **6**, 917–927 (1989).
- Weidinger, M., Varcoe, B. T. H., Heerlein, R. & Walther, H. Trapping states in the micromaser. *Phys. Rev. Lett.* **82**, 3795–3798 (1999).
- Maitre, X. *et al.* Quantum memory with a single photon in a cavity. *Phys. Rev. Lett.* **79**, 769–772 (1997).
- Englert, B. *et al.* Entangled atoms in micromaser physics. *Fortschr. Phys.* **46**, 897–926 (1998).
- Nogues, G. *et al.* Seeing a single photon without destroying it. *Nature* **400**, 239–242 (1999).
- Englert, B.-G., Sterpi, N. & Walther, H. Parity States in the one-atom maser. *Opt. Commun.* **100**, 526–535 (1993).
- Rempe, G. & Walther, H. Sub-poissonian atomic statistics in a micromaser. *Phys. Rev. A* **42**, 1650–1655 (1990).

- Rempe, G., Walther, H. & Klein, N. Observation of quantum collapse and revival in the one-atom maser. *Phys. Rev. A* **58**, 353–356 (1987).
- Raithel, G., Benson, O. & Walther, H. Atomic interferometry with the micromaser. *Phys. Rev. Lett.* **75**, 3446–3449 (1995).
- Benson, O., Raithel, G. & Walther, H. Quantum jumps of the micromaser field – dynamic behavior close to phase transition points. *Phys. Rev. Lett.* **72**, 3506–3509 (1994).
- Filipowicz, P., Javanainen, J. & Meystre, P. Theory of a microscopic maser. *J. Opt. Soc. B* **34**, 3077–3087 (1986).
- Cummings, F. W. & Rajagopal, A. K. Production of number states of the electromagnetic field. *Phys. Rev. A* **39**, 3414–3416 (1989).
- Jaynes, E. T. & Cummings, F. W. Quantum and semiclassical radiation theories. *Proc. IEEE* **51**, 89–109 (1963).
- Meystre, P., Rempe, G. & Walther, H. Very-low temperature behavior of a micromaser. *Opt. Lett.* **13**, 1078–1080 (1988).
- Krause, J., Scully, M. O. & Walther, H. State reduction and $|n\rangle$ -state preparation of a high-Q micromaser. *Phys. Rev. A* **36**, 4547–4550 (1987).
- Wehner, E., Seno, R., Sterpi, N., Englert, B.-G. & Walther, H. Atom pairs in the micromaser. *Opt. Commun.* **110**, 655–669 (1994).
- Brune, M. *et al.* Quantum Rabi oscillation: A direct test of field quantisation in a cavity. *Phys. Rev. Lett.* **76**, 1800–1803 (1996).

Acknowledgements

We thank B.-G. Englert for the theoretical calculations of the cavity photon number and for many useful discussions.

Correspondence and requests for materials should be addressed to H. W. (e-mail: herbert.walther@mpq.mpg.de).

Imaging the effects of individual zinc impurity atoms on superconductivity in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$

S. H. Pan[†], E. W. Hudson^{*}, K. M. Lang^{*}, H. Eisaki[‡], S. Uchida[‡] & J. C. Davis^{*}

^{*} Department of Physics, University of California, Berkeley, California 94720, USA

[‡] Department of Superconductivity, University of Tokyo, Tokyo 113-8656, Japan

[§] Department of Applied Physics, Stanford University, Stanford, California 94305, USA

Although the crystal structures of the copper oxide high-temperature superconductors are complex and diverse, they all contain some crystal planes consisting of only copper and oxygen atoms in a square lattice: superconductivity is believed to originate from strongly interacting electrons in these CuO_2 planes. Substituting a single impurity atom for a copper atom strongly perturbs the surrounding electronic environment and can therefore be used to probe high-temperature superconductivity at the atomic scale. This has provided the motivation for several experimental^{1–8} and theoretical studies^{9–20}. Scanning tunnelling microscopy (STM) is an ideal technique for the study of such effects at the atomic scale, as it has been used very successfully to probe individual impurity atoms in several other systems^{21–25}. Here we use STM to investigate the effects of individual zinc impurity atoms in the high-temperature superconductor $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. We find intense quasiparticle scattering resonances²⁶ at the Zn sites, coincident with strong suppression of superconductivity within ~ 15 Å of the scattering sites. Imaging of the spatial dependence of the quasiparticle density of states in the vicinity of the impurity atoms reveals the long-sought four-fold symmetric quasiparticle ‘cloud’ aligned with the nodes of the d -wave superconducting gap which is believed to characterize superconductivity in these materials.

[†] Present address: Department of Physics, Boston University, Boston, Massachusetts 02215, USA.

We have previously reported that STM spectroscopy of nominally undoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (BSCCO) crystals revealed atomic-scale regions having a relatively high density of low-energy quasiparticle states²⁶. The spatial and spectroscopic characteristics of these states were consistent with theories of quasiparticle scattering from atomic-scale impurities in a *d*-wave superconductor^{10,12,13}. However, the location in the crystal and the identity of these scattering centres was unknown. Furthermore, another STM experiment showed that introduction of atomic-scale defects onto the BSCCO surface also generated low-energy quasiparticle states²⁷. Recently, high-quality single crystals with controlled impurities have become available for STM studies. Here we describe the first use, to our knowledge, of STM to study local effects due to impurity atoms of known identity and crystal location in a high- T_c superconductor.

Our work focuses on Zn impurity atoms in BSCCO because they create strongly altered bulk superconducting properties^{1–4} and very

unusual spectroscopic effects^{5–8}. We use a home-built, low-temperature STM which has several unique features particularly suited to the study of high-temperature superconductors²⁶. The samples used are $\text{Bi}_2\text{Sr}_2\text{Ca}(\text{Cu}_{1-x}\text{Zn}_x)_2\text{O}_{8+\delta}$ single crystals grown by the floating zone method²⁸, with $x = 0.6\%$. They have been characterized to have a T_c of 84 K with a transition width of 4 K. These crystals are cleaved *in situ* at 4.2 K in cryogenic ultrahigh vacuum and are inserted into the STM immediately after cleavage.

A standard STM topographic image is first taken to determine the condition of the crystal surface, and a typical result is shown in Fig. 1a. A single crystal plane—believed to be usually the BiO layer with only the Bi atoms apparent in STM imaging—is exposed. The indistinguishability of these surfaces from those of non Zn-doped BSCCO^{26,27} is consistent with the fact that Zn dopant atoms do not reside at the BiO plane but rather at the Cu sites two layers below the surface.

To search for low-energy quasiparticle states associated with the Zn atoms, we next map the differential tunnelling conductance at zero sample bias, for a larger area of the surface. Because at any point the local density of states (DOS) is proportional to the differential conductance, this results in a map of the zero-energy DOS at the surface. Figure 1b is a typical DOS-map of a 500×500 Å region of the same surface as shown in Fig. 1a, and the overall dark background is indicative of a very low quasiparticle DOS near the Fermi level. This is as expected for a superconductor far below T_c . Remarkably, however, there are a number of randomly distributed bright sites corresponding to areas of high DOS, each with a distinct four-fold symmetric shape and the same relative orientation.

Further investigation of these bright sites reveals very significant differences between conductance spectra taken exactly at their centres and spectra taken at usual superconducting regions of the sample (dark areas in Fig. 1b). In Fig. 2 we show a comparison between two such spectra, from which several observations can be made. First, the spectrum at the centre of a bright site has a very strong intra-gap DOS peak, whose magnitude can be up to six times

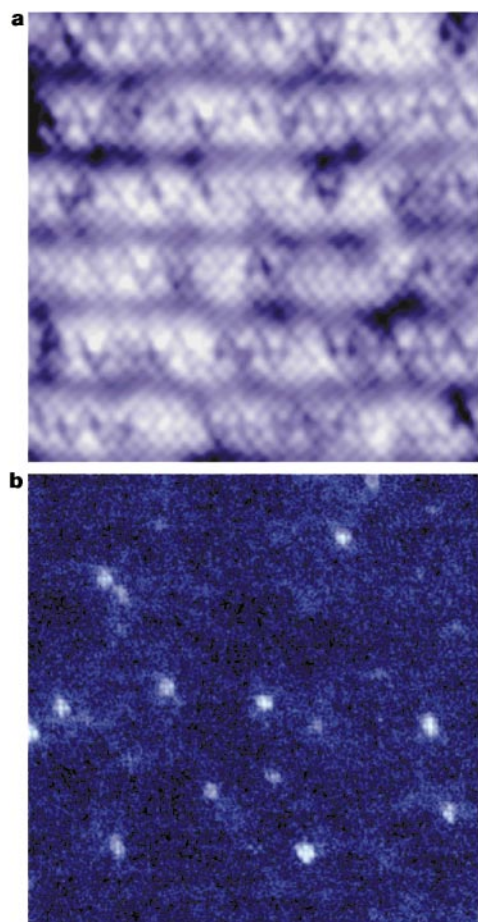


Figure 1 Topographic image and associated low-energy DOS-map of the surface layer (BiO) of a cleaved single crystal of BSCCO. **a**, Topographic image (150×150 Å) of the surface BiO plane which is exposed after cleavage of a Zn-BSCCO single crystal. The atoms are displaced from their ideal square lattice sites, forming a supermodulation along the crystal **b**-axis. The image of this surface is indistinguishable from those of non Zn-doped BSCCO. ($T = 4.2$ K, $I = 100$ pA, $V_{\text{sample}} = -100$ mV). **b**, A zero-bias differential tunnelling conductance map (500×500 Å), taken on a larger area of the same surface and with the same orientation as that shown in **a**. Because the differential tunnelling conductance dI/dV is proportional to the sample DOS, this is also a map of the quasiparticle DOS at the Fermi energy ($V_{\text{sample}} = 0$). Most of the image appears dark, which is indicative of a very low density of quasiparticles at the Fermi level. The Zn scattering sites appear as bright regions ~ 15 Å in extent, each clearly exhibiting a cross-shaped four-fold symmetric structure. Note that the orientations of all these cross-shaped features are the same. To acquire this image we operated at 4.2 K, set a 1 GΩ junction resistance ($I = 200$ pA, $V_{\text{sample}} = -200$ mV), and measure the tunnelling conductance with a standard low-frequency a.c. lock-in technique ($A_{\text{modulation}} = 500$ μV_{rms}, $f_{\text{modulation}} = 447.3$ Hz). The surface area shown in **a** is at the top centre of the DOS map in **b**.

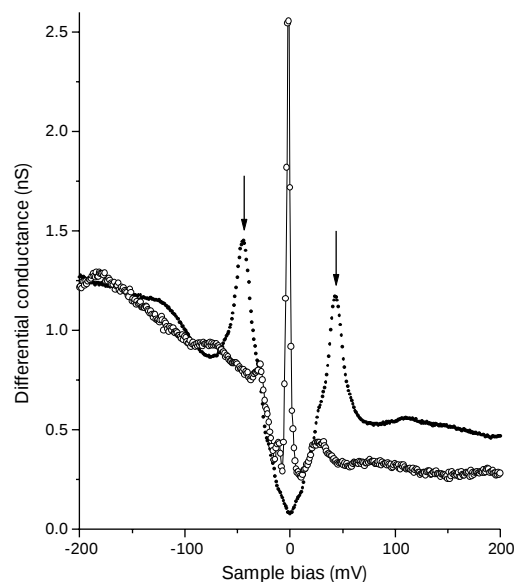


Figure 2 Two tunnelling spectra measured on the same surface shown in Fig. 1. Shown are differential tunnelling conductance spectra taken at two different locations on the Zn-BSCCO crystal. The spectrum of a 'usual' superconducting region of the sample, where Zn scatterers are absent (dark in Fig. 1b), is shown as filled circles. The superconducting coherence peaks are indicated by the arrows. The data shown as open circles, with an interpolating fine solid line, are the spectrum taken exactly at the centre of a bright scattering site. It shows both the intense scattering resonance peak centred at $\Omega = -1.5$ meV, and the very strong suppression of both the superconducting coherence peaks and gap magnitude at the Zn site.

greater than the normal-state conductance. Second, these peaks occur at typical values of quasiparticle energy (as measured from the Fermi level) of $\Omega = -1.5 \pm 0.5$ meV. Third, at these sites the superconducting coherence peaks (identified by the arrows in Fig. 2) are strongly suppressed, indicating the almost complete destruction of superconductivity. All of these phenomena are among the theoretically predicted characteristics of a very strong quasiparticle scattering resonance at a single impurity atom in a *d*-wave

superconductor^{12,13,15,16,18,20}.

To identify the source of these resonances, we next acquire simultaneous pairs of high-resolution topographic images and resonant-energy DOS-maps, centred on the resonance sites. Comparison between each topograph/DOS-map pair shows that the centre of a scattering resonance always coincides with the site of a surface Bi atom, as shown for example in Fig. 3a and b. From the known crystal structure of BSCCO, a Cu site where the Zn dopant atom could reside is directly below each exposed Bi atom, and separated from it by the intervening SrO₂ layer. Furthermore, the characteristic spatial features of the resonance (as shown for example in Fig. 3b) have never been observed in non-Zn-doped BSCCO crystals. Finally, the density of the observed scattering resonance sites is about $x = 0.2\%$, in reasonable agreement with the nominal Zn doping concentration, and no population of weak scatterers similar to those in ref. 26 is observed in these Zn-BSCCO crystals. We therefore attribute the observed very strong resonances to quasiparticle scattering at the Zn impurity atoms.

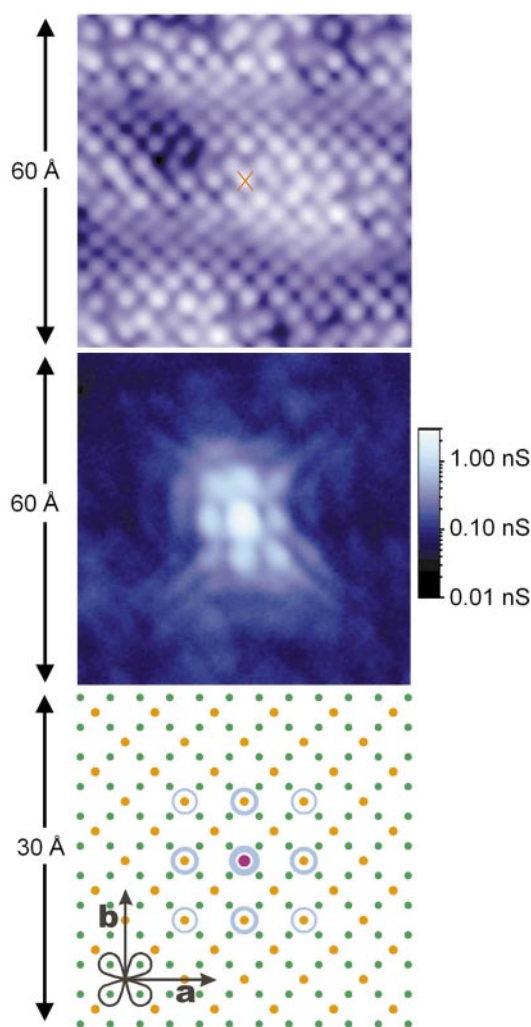


Figure 3 Relationship between the position of the Bi atoms on the crystal surface, the resonant DOS structure at the Zn atom, and the position of the Cu and O atoms in the superconducting plane two layers below. **a** and **b** are simultaneously acquired, 60×60 Å, high-spatial-resolution images of the topography and the differential conductance at $V_{\text{sample}} = -1.5$ mV. The bright centre of the scattering resonance in **b** coincides with the position of the Bi atom marked by an X in **a**. Note that **b** uses a logarithmic scale of intensity as the features at largest distances from the Zn site are more than two orders of magnitude weaker than the resonance peak. The inner bright cross is oriented with the nodes of the *d*-wave gap (as indicated schematically in **c**). The weaker outer features, including the ~ 30 -Å-long “quasiparticle beams” at 45° to the inner cross, are oriented with the gap maxima. **c**, A 30×30 Å schematic representation of the square CuO₂ lattice, which is located two layers below the exposed surface BiO plane, showing its relative orientation to the BiO surface in **a**. The Cu atoms are indicated as orange circles and the O atoms as green circles, while the Zn scatterer (in purple) is at the central point. The crystal **a**-axis and **b**-axis orientations, and the orientations of the maxima and nodes of the *d*-wave superconducting gap, as measured by ARPES experiments, are also shown. The location of the maxima in the DOS, as measured from **b**, are shown schematically as open blue circles surrounding the atomic sites with which they appear to be associated. The intensity of the DOS at these sites is shown schematically by the thickness of the line forming these circles.

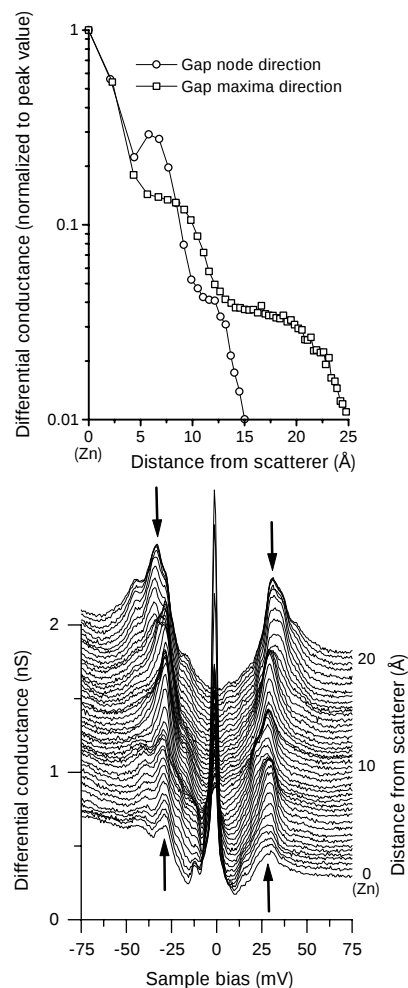


Figure 4 Evolution of the differential tunnelling conductance at the resonance energy, and the full conductance spectrum, with distance from the Zn site. **a**, Differential tunnelling conductance (normalized to the peak value) versus distance r from the centre of a Zn scatterer, along the **a**-axis or **b**-axis (gap-node) directions (shown as open circles) and along the Cu–O bond (gap-maxima) directions (shown as open squares). Each data point is the local DOS at distance r averaged over an inclusive angle of 10° around the given directions. **b**, Waterfall plot of a series of tunnelling spectra taken on a line from the scatterer along the **b**-axis direction in 0.5 -Å steps. This clearly shows the suppression of the coherence peaks (whose absence at the Zn scatterer is indicated by upward pointing arrows), and recovery of the superconductivity (along with the coherence peaks whose presence is indicated by downward pointing arrows) within a distance of ~ 15 Å from the scattering site.

The DOS-maps at Zn scattering resonance sites show complex structures consisting of spatial elements whose DOS varies over two orders of magnitude. To show these elements in full detail, Fig. 3b is presented in a logarithmic intensity scale. Many features are apparent in this representation. In addition to the central bright spot, we see that a dominant feature is a relatively bright cross-shaped region aligned along the crystal **a**- and **b**-axes, and extending to a radius of about 10 Å. Concentric with this bright cross, and rotated 45° degrees relative to it, another cross-shaped feature with considerably lower intensity extends to approximately 30 Å from the centre.

The quasiparticle DOS at the resonance energy does not decay monotonically with distance r from the scatterer but rather oscillates, producing local minima and maxima in the DOS-map. Analysis of the correspondence of DOS local maxima in Fig. 3b, with the simultaneously imaged atomic positions in Fig. 3a, indicates that they coincide with positions of only some of the atomic sites near the scattering centre. To help clarify the spatial register of these DOS oscillations to the crystal, we show schematically in Fig. 3c the positions of the Cu and O atoms in the CuO₂ plane with the positions of DOS local maxima at the BiO plane overlaid as blue circles. Also shown are the orientations of the crystal **a**- and **b**-axes, and that of the d -wave gap nodes as determined by angle-resolved photoemission spectroscopy (ARPES)²⁹. The positions of the four nearest-neighbour Cu atoms to the Zn have no DOS local maxima associated with them. On the other hand, the positions of the eight second-nearest-neighbour and third-nearest-neighbour Cu atoms coincide with local maxima in the DOS image, and appear to form a 'box' around the scatterer (Fig. 3b).

In Fig. 4a we plot the normalized DOS at the resonance energy as a function of r along both the gap-node and gap-maxima directions. Due to the strength of the oscillations it is difficult to identify, for comparison with theory^{12,13,19,20}, a power law governing the fall-off of the DOS with r in these directions. However, the rate of decay is clearly faster in the gap-node directions, while towards the gap maxima weak "quasiparticle beams" are apparent up to 30 Å from the scatterer.

The Zn impurity atom locally affects not only the low-energy DOS, but also features at the gap energy scale. Its effect on the magnitude of the coherence peaks, and on the width of the superconducting gap, is shown in Fig. 4b, which is a plot of the measured evolution of the complete DOS spectrum as the tip of the STM is moved away from the scattering site along the **b**-axis. The superconducting coherence peaks are strongly suppressed at the Zn scattering site, and recover to their usual value over a distance of about 15 Å.

We have found that all the above phenomena are characteristic of large numbers of scattering sites in several Zn-doped BSCCO crystal samples. When compared to the previous observations at the weaker scatterers of unknown identity²⁶, the Zn resonances are clearly much stronger, which in turn allows us to determine their detailed spatial structure with far greater resolution.

Some observations at the Zn scatterers are in excellent agreement with theoretical predictions. For example, theoretical proposals that quasiparticle scattering resonances, and localized impurity states, could be created by individual impurity atoms in a d -wave superconductor^{10–13,15–20} are here shown to be correct. The strength of the scattering leading to these resonances, in terms of a phase shift δ_0 , may be calculated, in the context of refs 12 and 13, using the ratio of the resonance energy Ω to the energy gap Δ_0 :

$$\left| \frac{\Omega}{\Delta_0} \right| \approx \frac{\pi c/2}{\ln(8/\pi c)} \quad (1)$$

where $c = \cot(\delta_0)$. Using an average gap value (away from the impurities) of $\Delta_0 = 45$ mV and $\Omega = -1.5$ mV, we obtain a phase shift of 0.48π . This confirms that scattering from Zn is very close to the unitary limit ($\delta_0 = \pi/2$)^{9–20}. The predicted particle-hole symmetry breaking at a non-magnetic impurity atom scattering resonance^{12,13,20} is

directly confirmed. The predicted four-fold symmetry in the spatial distribution of the DOS near an impurity atom^{10,12,13,20}, and its alignment with the d -wave gap nodes, are also clearly observed.

Our measurements also provide microscopic validation for models proposed to explain the results of several previous experiments. A proposal that superfluid density reductions can be explained by non-superconducting regions of area $\pi\xi^2$ where ξ is the coherence length around each impurity atom³ (the 'Swiss cheese' model) is directly validated, both qualitatively and quantitatively, by the suppression of superconductivity near the Zn site shown in Fig. 4b. The strong non-zero energy quasiparticle DOS at the impurity atoms, which was postulated to explain infrared spectroscopy results⁶, is also directly confirmed by the results shown in Fig. 2.

Despite this consistency between previous theoretical and experimental results and our measurements, we have also observed several unexpected phenomena. These include the extended "quasiparticle beams" oriented with the gap maxima, and the local maxima in the DOS associated with specific atomic sites. These phenomena suggest that to fully explain our results, a more sophisticated model is needed, which considers effects such as the symmetry of the interlayer tunnelling matrix element ($t_\perp$) (ref. 30), the band structure^{17,19}, or local magnetic moments at the Zn sites^{31,32}.

For example, the angular dependence of the interlayer tunnelling matrix element, which connects the BiO surface layer to the CuO₂ plane, may partially explain one of the unexpected observations. This matrix element is believed to have the same symmetry as the $d_{x^2-y^2}$ order-parameter³⁰, and thus would reduce the tunnelling signal for states along the gap-node directions and enhance it for states towards the gap-maxima. This would result in an apparent rapid decay of the DOS in the gap-node directions and an apparent long extension of states oriented towards the gap maxima, which is consistent with our observations. In fact, within this framework, the clear observation at the surface of the cross-shaped DOS towards the gap nodes indicates that this characteristic feature would be even more dominant in the CuO₂ plane. However, the register of the DOS to specific atoms cannot be explained by effects of $t_\perp$. In particular, no DOS maxima are associated with the nearest-neighbour Cu sites although $t_\perp$ would apparently allow local maxima at these locations.

Finally, the reported quasiparticle scattering spectroscopy and DOS imaging at individual impurity atoms in the copper oxides not only reveals new information on electronic impurity states in high-temperature superconductors but also opens a new avenue for research into these important materials. □

Received 1 September; accepted 3 December 1999.

- Maeda, A., Yabe, T., Takebayashi, S., Hase, M. & Uchinokura, K. Substitution of 3d metals for Cu in Bi₂(Sr_{0.6}Ca_{0.4})₂Cu₂O₇. *Phys. Rev. B* **41**, 4112–4117 (1990).
- Ishida, K. *et al.* Gapless superconductivity in Zn-doped YBa₂Cu₃O₇ studied by Cu NMR and NQR. Possibility of d -wave superconductivity in high- T_c oxides *Physica C* **179**, 29–38 (1991).
- Bonn, D. A. *et al.* Comparison of the influence of Ni and Zn impurities on the electromagnetic properties of YBa₂Cu₃O_{8.95}. *Phys. Rev. B* **50**, 4051–4063 (1994).
- Fukuzumi, Y., Mizuhashi, K., Takenaka, K. & Uchida, S. Universal superconductor-insulator transition and T_c depression in Zn-substituted high- T_c cuprates in the underdoped regime. *Phys. Rev. Lett.* **76**, 684–687 (1996).
- Nachumi, B. *et al.* Muon spin relaxation studies of Zn-substitution effects in high- T_c cuprate superconductors *Phys. Rev. Lett.* **77**, 5421–5424 (1996).
- Basov, D.N., Dabrowski, B. & Timusk, T. Infrared probe of transition from superconductor to nonmetal in YBa₂(Cu_{1-x}Zn_x)₂O₈. *Phys. Rev. Lett.* **81**, 2132–2135 (1998).
- White, P.J. *et al.* Zn impurities in Bi₂Sr₂Ca(Cu_{1-x}Zn_x)₂O_{8.66} – electronic structure evolution. Preprint cond-mat/9901349 at (<http://xxx.lanl.gov>) (1999).
- Fong, H.F. *et al.* Effect of nonmagnetic impurities on the magnetic resonance peak in YBa₂Cu₃O₇. *Phys. Rev. Lett.* **82**, 1939–1942 (1999).
- Lee, P. A. Localized states in a d -wave superconductor *Phys. Rev. Lett.* **71**, 1887–1890 (1993).
- Byers, J.M., Flatté, M.E. & Scalapino, D.J. Influence of gap extrema on the tunneling conductance near an impurity in an anisotropic superconductor *Phys. Rev. Lett.* **71**, 3363–3366 (1993).
- Sun, Y. & Maki, K. Impurity effects in d -wave superconductors *Phys. Rev. B* **51**, 6059–6063 (1995).
- Balatsky, A.V., Salkola, M.I. & Rosengren, A. Impurity-induced virtual bound states in d -wave superconductors *Phys. Rev. B* **51**, 15547–15551 (1995).
- Salkola, M.I., Balatsky, A.V. & Scalapino, D.J. Theory of scanning tunneling microscopy probe of impurity states in a d -wave superconductor *Phys. Rev. Lett.* **77**, 1841–1844 (1996).
- Hirschfeld, P.J. & Puttিকা, W.O. Theory of thermal conductivity in YBa₂Cu₃O_{7-δ}. *Phys. Rev. Lett.* **77**, 3909–3912 (1996).

15. Franz, M., Kallin, C. & Berlinsky, A.J. Impurity scattering and localization in *d*-wave superconductors *Phys. Rev. B* **54**, R6897–R6900 (1996).
16. Salkola, M.I. & Schrieffer, J.R. Unusual states of inhomogeneous $d_{x^2-y^2} + id_{xy}$ superconductors *Phys. Rev. B* **58**, R5952–R5955 (1998).
17. Atkinson, W.A. & MacDonald, A.H. Visualizing quasiparticle scattering resonances *Science* **285**, 57–58 (1999).
18. Tsuchiura, H., Tanaka, Y., Ogata, M. & Kashiwaya, S. Quasiparticle properties around a nonmagnetic impurity in the superconducting state of the two-dimensional *t*-*J* model. *J. Phys. Soc. Jpn* **68**, 2510–2513 (1999).
19. Flatté, M.E. & Byers, J.M. Local electronic structure of defects in superconductors *Solid State Phys.* **52**, 137–228 (1999).
20. Salkola, M.I., Balatsky, A.V. & Schrieffer, J.R. Spectral properties of quasiparticle excitations induced by magnetic moments in superconductors *Phys. Rev. B* **55**, 12648–12661 (1997).
21. Crommie, M.F., Lutz, C.P. & Eigler, D.M. Imaging standing waves in a two-dimensional gas *Nature* **363**, 524–527 (1993).
22. Zheng, J.F. et al. Scanning tunneling microscopy studies of Si donors (Si_{Ga}) in GaAs. *Phys. Rev. Lett.* **72**, 1490–1493 (1994).
23. Yazdani, A., Jones, B.A., Lutz, C.P., Crommie, M.F. & Eigler, D.M. Probing the local effects of magnetic impurities on superconductivity *Science* **275**, 1767–1770 (1997).
24. Madhavan, V., Chen, W., Jamneala, T., Crommie, M.F. & Wingreen, N.S. Tunneling into a single magnetic atom: spectroscopic evidence of the Kondo resonance *Science* **280**, 567–569 (1998).
25. Wittneven, Chr., Dombrowski, R., Morgenstern, M. & Wiesendanger, R. Scattering states of ionized dopants probed by low temperature scanning tunneling spectroscopy *Phys. Rev. Lett.* **81**, 5616–5619 (1998).
26. Hudson, E.W., Pan, S.H., Gupta, A.K., Ng, K.-W. & Davis, J.C. Atomic-scale quasiparticle scattering resonances in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ *Science* **285**, 88–91 (1999).
27. Yazdani, A., Howald, C.M., Lutz, C.P., Kapitulnik, A. & Eigler, D.M. Impurity-induced bound excitations on the surface of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ *Phys. Rev. Lett.* **83**, 176–179 (1999).
28. Motohira, N., Kuwahara, K., Hasegawa, T., Kishio, K. & Kitazawa, K. Single crystal growth of $\text{Bi}_2\text{Sr}_2\text{Ca}_{n-1}\text{Cu}_n\text{O}_y$ superconductors by the floating zone method *J. Ceram. Soc. Jpn* **97**, 1009–1014 (1989).
29. Wells, B.O. et al. Evidence for *k*-dependent, in-plane anisotropy of the superconducting gap in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ *Phys. Rev. B* **46**, 11830–11834 (1992).
30. Andersen, O. K., Jepsen, O., Liechtenstein, A. I. & Mazin, I. I. Plane dimpling and saddle-point bifurcation in the band structures of optimally doped high-temperature superconductors: a tight-binding model. *Phys. Rev. B* **49**, 4145–4157 (1994).
31. Mahajan, A.V., Alloul, H., Collin, G. & Marucco, J.F. ^{69}Y NMR probe of Zn induced local moments in $\text{YBa}_2(\text{Cu}_{1-x}\text{Zn}_x)_2\text{O}_{6+x}$ *Phys. Rev. Lett.* **72**, 3100–3103 (1994).
32. Nagaosa, N. & Lee, P.A. Kondo effect in high- T_c cuprates. *Phys. Rev. Lett.* **79**, 3755–3758 (1997).

Acknowledgements

We thank A. Balatsky, D. Bonn, M. Crommie, M. Flatté, M. Franz, S. Kashiwaya, A. de Lozanne, A. MacDonald, V. Madhavan, M. Ogata, J. Orenstein, D. J. Scalapino, Z.-X. Shen, Y. Tanaka and D. van der Marel for conversations and communications. This work was supported by the LDRD Program of the Lawrence Berkeley National Laboratory under contract to the Department of Energy, by the D. & L. Packard Foundation, by an IBM predoctoral fellowship (K.M.L.), by Grant-in-Aid for Scientific Research on Priority Area (Japan), and by a COE grant from the Ministry of Education, Japan.

Correspondence and requests for materials should be addressed to J.C.D. (e-mail: jcdavis@socrates.berkeley.edu).

High-efficiency fluorescent organic light-emitting devices using a phosphorescent sensitizer

M. A. Baldo*, M. E. Thompson† & S. R. Forrest*

* Center for Photonics and Optoelectronic Materials (POEM), Department of Electrical Engineering and the Princeton Materials Institute, Princeton University, Princeton, New Jersey 08544, USA

† Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA

To obtain the maximum luminous efficiency from an organic material, it is necessary to harness both the spin-symmetric and anti-symmetric molecular excitations (bound electron–hole pairs, or excitons) that result from electrical pumping. This is possible if the material is phosphorescent, and high efficiencies have been observed in phosphorescent^{1,2} organic light-emitting devices³. However, phosphorescence in organic molecules is rare at room temperature. The alternative radiative process of fluor-

escence is more common, but it is approximately 75% less efficient, due to the requirement of spin-symmetry conservation⁴. Here, we demonstrate that this deficiency can be overcome by using a phosphorescent sensitizer to excite a fluorescent dye. The mechanism for energetic coupling between phosphorescent and fluorescent molecular species is a long-range, non-radiative energy transfer: the internal efficiency of fluorescence can be as high as 100%. As an example, we use this approach to nearly quadruple the efficiency of a fluorescent red organic light-emitting device.

Light is generated in organic materials from the decay of molecular excited states, also known as excitons. Understanding the properties and interactions of excitons is crucial to the design of efficient organic devices for use in displays, lasers and other illumination applications. For example, the spin-symmetry of an exciton determines its probability of radiative recombination and also its multiplicity. Spin-symmetric excitons with a total spin of $S = 1$ have a multiplicity of three and are known as triplets. Spin-antisymmetric excitons ($S = 0$) have a multiplicity of one and are known as singlets. During electrical excitation approximately one singlet exciton is created for every three triplet excitons⁴, but because the ground state is typically also spin-antisymmetric, only relaxations of singlet excitons conserve spin and generate fluorescence. Usually the energy in triplet excitons is wasted; however, given some perturbation in symmetry, triplets may slowly radiatively decay, producing the delayed luminescence known as phosphorescence. Although it is often inefficient, phosphorescence may be enhanced if spin–orbit coupling mixes the singlet and triplet states, an effect often promoted by the presence of a heavy metal atom⁵. Indeed, phosphorescent dyes with these properties have demonstrated very high-efficiency electroluminescence^{1,2}.

Very few organic materials have been found to be capable of efficient room-temperature phosphorescence from triplets^{1,2,6}. In contrast, many organic molecules exhibit fluorescence^{7,8} and fluorescence is also unaffected by triplet–triplet annihilation, which degrades phosphorescent emission efficiency at high excitation densities¹. Consequently, fluorescent materials are suited to many electroluminescent applications, particularly those such as passive matrix displays that require high excitation densities.

It is desirable therefore to find a process whereby triplets that are formed after electrical excitation are not wasted, but are instead transferred to the singlet excited state of a fluorescent dye. There are two mechanisms for triplet–singlet energy transfer from a donor molecule (D) to an acceptor (A). In Dexter transport⁵, the exciton hops directly between molecules. This is a short-range process dependent on the overlap of molecular orbitals of neighbouring molecules. It also preserves the symmetry of the donor and acceptor pair⁵. Thus, a triplet–singlet energy transfer is not possible by a Dexter mechanism. A change in spin-symmetry is possible if the donor exciton breaks up and reforms on the acceptor by incoherent electron exchange⁵. However, this process is considered to be relatively unlikely as it requires the dissociation of the donor exciton, which in most molecular systems has a binding energy of ~ 1 eV.

The alternative mechanism is Förster energy transfer⁵. Here, molecular transition dipoles couple and exchange energy. The efficiency of energy transfer (η_{ET}) is:

$$\eta_{\text{ET}} = \frac{k_{\text{ET}}}{k_{\text{ET}} + k_{\text{r}} + k_{\text{nr}}} \quad (1)$$

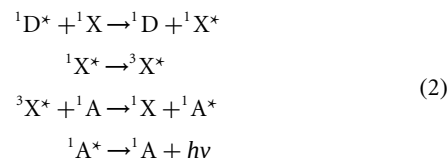
Here k_{ET} is the rate of Förster energy transfer from D to A and k_{r} and k_{nr} are the radiative and non-radiative rates on the donor, respectively. From equation (1), energy transfer is efficient if $k_{\text{ET}} > k_{\text{r}} + k_{\text{nr}}$; however, in Förster's theory k_{ET} is proportional to the oscillator strength of the donor transition⁵, as is k_{r} . Thus, η_{ET} is approximately independent of oscillator strength if $k_{\text{r}} \gg k_{\text{nr}}$, that is, if the donor is efficiently phosphorescent then it is possible to

obtain triplet–singlet energy transfer by the Förster mechanism. We note that a ‘pure’ triplet has an infinite lifetime and no probability of Förster energy transfer because the oscillator strength of its decay is zero. However, the slightest perturbation (that is, some overlap between the triplet and ground state on the donor) can counteract the presence of non-radiative modes and yield efficient energy transfer. Such triplet–singlet energy transfers were predicted by Förster⁹ and confirmed by Ermolaev and Sveshnikova¹⁰, who detected the energy transfer using a range of phosphorescent donors and fluorescent acceptors in rigid media at 77 K or 90 K. Large transfer distances were observed; for example with triphenylamine as the donor and chrysoidine as the acceptor, the interaction range is 52 Å.

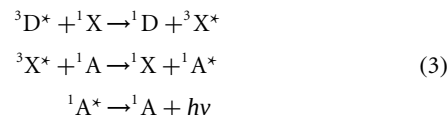
Unfortunately, when a fluorescent acceptor is doped directly into a phosphorescent donor material, the close proximity of the donor and acceptor increases the likelihood of Dexter transfer between the donor and the acceptor triplets. Once excitons reach the acceptor triplet state, they are effectively lost since these fluorescent dyes typically exhibit extremely inefficient phosphorescence (that is, $k_{\text{NR}} \gg k_{\text{R}}$).

Another technique is to dope both the phosphorescent material and the fluorescent acceptor into a conductive organic host. Ideally, the phosphor then sensitizes the energy transfer from the host, now acting as the donor, to the fluorescent acceptor. Cascade Förster energy transfer of singlets has been demonstrated for fluorescent materials¹¹; however, here all energy is ideally transferred into the

triplet state of the sensitizer, where it is then transferred to the singlet state of the fluorescent dye, that is:



and



Here, the photon energy is $h\nu$, and the donor, sensitizer and fluorescent acceptor are represented by D, X and A, respectively. Triplet and singlet states are signified by a superscript 3 or 1, respectively, and excited states are marked by asterisks. The multi-stage energy transfer is described schematically in Fig. 1. Dexter transfers are indicated by dotted arrows and Förster transfers by solid arrows. Processes resulting in a loss in efficiency are marked with a cross. In addition to the energy transfer paths shown in Fig. 1, direct electron–hole recombination is possible on the phosphorescent and fluorescent dopants as well as the host. Triplet exciton formation after charge recombination on the fluorescent dye is another potential loss mechanism.

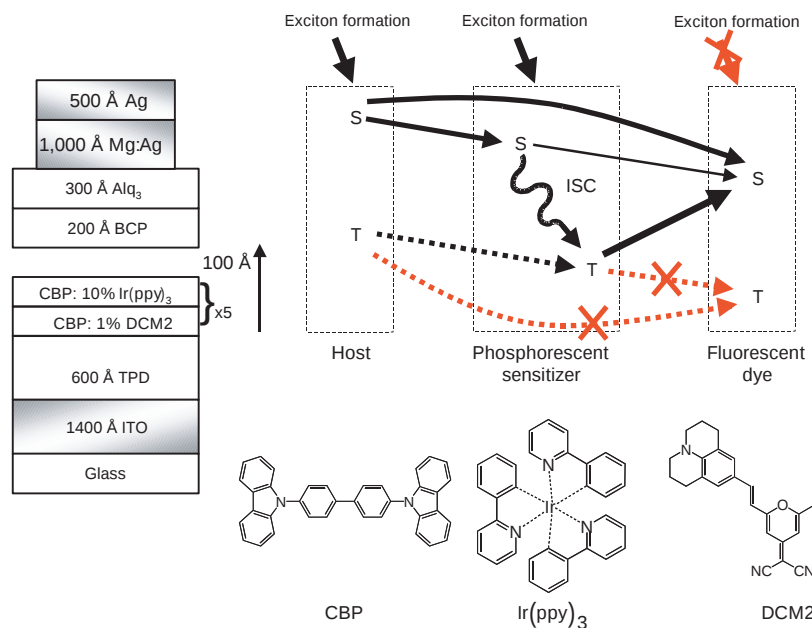


Figure 1 Proposed energy transfer mechanisms in the sensitized system. In principle, all excitons are transferred to the singlet state of the fluorescent dye, as triplets in the dye non-radiatively recombine. Förster transfers are represented by solid lines and Dexter transfers by dotted lines. Electron–hole recombination creates singlet (S) and triplet (T) excitons in the host material, although as indicated, charge trapping may be responsible for exciton formation in the other materials as well. There is a probability of direct transfer into the singlet state of the fluorescent dye by a Förster process, or Dexter transfer into the triplet state. This is a source of loss and is indicated by a cross. Singlet excitons in the phosphor are subject to intersystem crossing (ISC) and transfer to the triplet state. From this state, the triplets may either dipole–dipole couple with the singlet state of the fluorescent dye or, in another loss mechanism, they may Dexter transfer to the triplet state. Direct formation of triplets on the fluorescent dye is an additional path to loss. Inset, the structure of the principal materials and electroluminescent devices fabricated in this work. We use the green emitting phosphor² *fac*-tris(2-phenylpyridine) iridium ($\text{Ir}(\text{ppy})_3$) and the red fluorescent dye¹⁷ [2-methyl-6-[2-(2,3,6,7-tetrahydro-1H,5H-benzo[*ij*]quinolizin-9-yl) ethenyl]-4H-pyran-4-ylidene] propane-dinitrile (DCM2). DCM2 absorbs in the green

and it emits at wavelengths between $\lambda = 570\text{nm}$ and $\lambda = 650\text{nm}$ (ref. 14). To demonstrate the multiple-stage transfer, we used 4,4'-N,N'-dicarbazole-biphenyl (CBP) as the donor and host material¹⁸. Organic layers were deposited by high vacuum (10^{-6} Torr) thermal evaporation onto a clean glass substrate pre-coated with a layer 1,400 Å in thickness of transparent and conductive indium tin oxide. A layer 600 Å in thickness of N,N'-diphenyl-N,N'-bis(3-methylphenyl)-[1,1'-biphenyl]-4,4'-diamine (TPD) is used for hole transport. We used an alternating series of layers 10 Å in thickness of 10% $\text{Ir}(\text{ppy})_3$ /CBP and 1% DCM2/CBP. In total, 10 doped layers were grown with a total thickness of 100 Å. Excitons were confined within the luminescent region by a layer 200 Å in thickness of the wide-energy-gap material 2,9-dimethyl-4,7-diphenyl-1,10-phenanthroline (bathocuproine, or BCP). A layer 300 Å in thickness of tris-(8-hydroxyquinoline) aluminium (Alq_3) is used to transport electrons to the luminescent region and to reduce absorption at the cathode. Metal deposition through a shadow mask with openings 1 mm in diameter defined cathodes consisting of a layer 1000 Å in thickness of 25:1 Mg/Ag, with an Ag cap 500 Å in thickness.

The structure of the organic devices and the principal materials used to demonstrate the sensitized energy transfer are shown in the inset of Fig. 1. We used CBP as the host and exciton donor, the green phosphor Ir(ppy)₃ as the sensitizer and the red fluorescent dye DCM2 as the acceptor. As a control, two other organic light-emitting device (OLED) structures were made. In the first, Ir(ppy)₃ was replaced by Alq₃, which has similar emission and absorption spectra, but no observable phosphorescence at room temperature. From the spectral overlap with DCM2 and the photoluminescent efficiencies^{12,13} of Alq₃ and Ir(ppy)₃, the Förster transfer radii for Alq₃ to DCM2 and Ir(ppy)₃ to DCM2 are both calculated to be ~ 40 Å (ref. 14). In the second device, the sensitizer was completely omitted in order to examine direct energy transfer from the donor, CBP, to the acceptor, DCM2.

The external quantum efficiency (photons per electron) of the DCM2 portion of the emission spectrum of these three devices is shown as a function of injected current density in Fig. 2. The DCM2 emission efficiency of the device containing the phosphorescent sensitizer is significantly higher than its fluorescent analogue. Indeed, the peak efficiency of (3.360.1)% is significantly higher than the best result of $\sim 2\%$ observed for DCM2-based OLEDs in

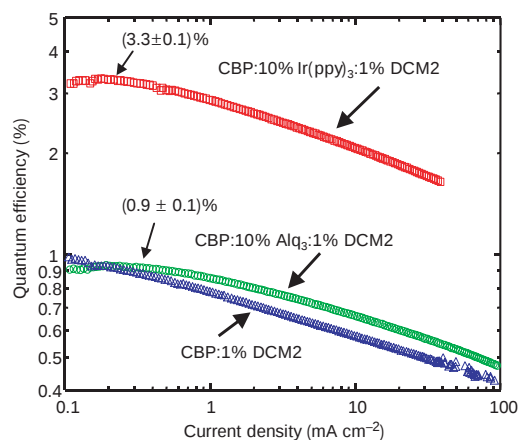


Figure 2 The external quantum efficiencies of DCM2 emission in the three devices. The sensitizing action of Ir(ppy)₃ improves the efficiency. We also note that the presence of Alq₃ in the all-fluorescent devices makes little or no difference. The CBP/10% Ir(ppy)₃/1% DCM2 device reaches a brightness of 100 cd m⁻² at 9.9 V (3 lm W⁻¹).

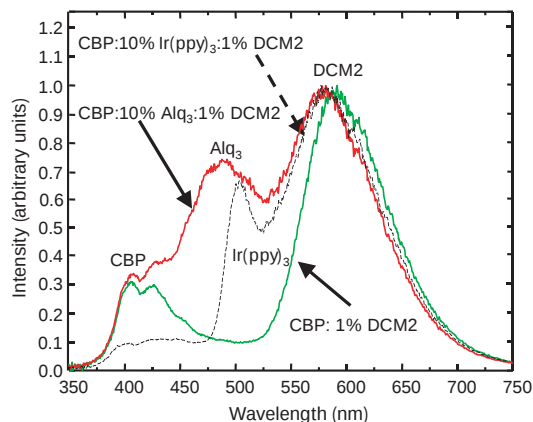


Figure 3 The spectra of the three electroluminescent devices fabricated in this work. Characteristic peaks are observed for CBP (at wavelength $\lambda \approx 400$ nm), TPD ($\lambda \approx 430$ nm), Alq₃ ($\lambda \approx 490$ nm), Ir(ppy)₃ ($\lambda \approx 500$ nm) and DCM2 ($\lambda \approx 590$ nm). Approximately 80% of the photons in the Ir(ppy)₃ device are emitted by DCM2. All spectra were recorded at a current density of ~ 1 mA cm⁻². Note that the peak intensities of each of the spectra are normalized for comparison.

previous studies¹⁵, suggesting that host triplets are transferred to the fluorescent singlet state. We also note that addition of Alq₃ to the unsensitized CBP/DCM2 device makes no significant difference to the maximum efficiency of (0.960.1)% of DCM2 emission.

The emission spectra of the OLEDs are shown in Fig. 3. All devices show energy transfer to the fluorescent dye. By calculating the area under the various spectral peaks, we find that approximately 80% of photons are emitted by DCM2 in the device containing the Ir(ppy)₃ sensitizer. The remainder contribute to CBP luminescence at $\lambda \approx 400$ nm, TPD luminescence at $\lambda \approx 430$ nm and Ir(ppy)₃ luminescence at $\lambda \approx 500$ nm. In the device doped with 10% Alq₃, an emission peak is also observed at $\lambda \approx 490$ nm. This is consistent with earlier observations¹⁶ of Alq₃ emission in a non-polar host such as CBP.

Conclusive evidence of the energy transfer of equations (2) and (3) is provided by examining the transient behaviour of the DCM2 and Ir(ppy)₃ components of the emission spectra. These data are shown in Fig. 4, and were obtained by applying a ~ 100 ns electrical pulse to the electroluminescent device. The resulting emission was measured using a streak camera. If a fraction of the DCM2 emission originates by transfer from Ir(ppy)₃ triplets, then the proposed energy transfer must yield delayed DCM2 fluorescence. Furthermore, since the radiative lifetime of DCM2 is much shorter than that of Ir(ppy)₃, the transient decay of DCM2 should match that of Ir(ppy)₃. After an initial peak, most probably due to CBP/DCM2 singlet-singlet transfer and exciton formation on DCM2, the DCM2 decay does indeed trace the Ir(ppy)₃ decay. The transient lifetime of Ir(ppy)₃ in this system is ~ 100 ns, compared to a lifetime of ~ 500 ns in the absence of DCM2. Since these lifetimes are inversely proportional to the rates of exciton transfer and relaxation, respectively, this confirms an energy transfer of $\sim 80\%$. The decrease in the triplet lifetime as a result of energy transfer to the fluorescent acceptor is advantageous. Not only does it increase the transient response of the system, but it should also reduce the probability of triplet-triplet annihilation. Thus, it is expected that this multi-stage energy transfer will reduce the quenching of triplet states, thereby further enhancing the potential efficiency of sensitized fluorescence.

The transient response of the sensitized system may be further examined to determine the relative efficiencies of the singlet-to-singlet and triplet-to-singlet energy transfer pathways. By subtracting

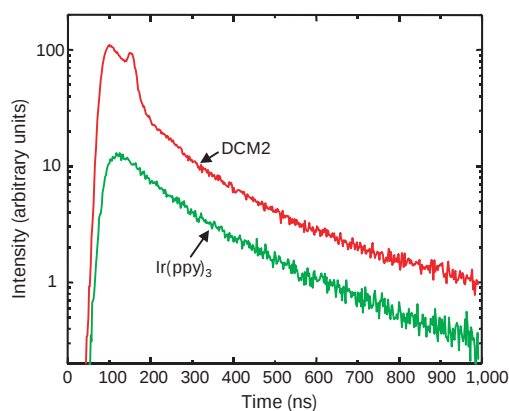


Figure 4 The transient response of the DCM2 and Ir(ppy)₃ spectral components in the CBP/10% Ir(ppy)₃/1% DCM2 device. The transient lifetime of DCM2 is ~ 1 ns, thus in the case of energy transfer from Ir(ppy)₃, the response of DCM2 should be governed by the transient lifetime of Ir(ppy)₃. After the initial 100-ns-wide electrical excitation pulse, this is clearly the case, demonstrating that energy is transferred from the triplet state in Ir(ppy)₃ to the singlet state in DCM2. However, during the excitation pulse, singlet transfer to DCM2 is observed, resulting in the ripples in the transient response. These ripples are due to fluctuations in the current density and the discharge of traps at the falling edge of the pulse. We note that the trends in the DCM2 and Ir(ppy)₃ transient responses eventually diverge. This is caused by charge trapped on DCM2 molecules recombining and causing luminescence.

the transient decay of Ir(ppy)₃ from the DCM2 transient, the ratio of instantaneous to delayed DCM2 fluorescence is found to be ~ 1:2. Consequently, we expect that the sensitized OLED should have a quantum efficiency roughly triple that of an OLED that does not exhibit delayed fluorescence. Indeed, from the data in Fig. 2, the improvement in peak quantum efficiencies is (3.7 ± 0.5)%. Ideally the ratio of instantaneous to delayed fluorescence should reflect the ratio of singlets to triplets (1:3), or even be weighted towards triplets if intersystem crossing on Ir(ppy)₃ is significant. It is likely that direct triplet exciton formation on DCM2 contributes to some loss. In addition, the likelihood of triplet transfer from CBP to DCM2, CBP to Ir(ppy)₃, or even Ir(ppy)₃ to CBP is unknown, so it is possible that some triplets are transferred directly to DCM2, bypassing the sensitizer.

In conclusion, we have demonstrated a general technique for improving the efficiency of fluorescence in OLEDs. Further improvement in energy transfer efficiency could be expected by mixing the host, phosphorescent sensitizer and fluorescent dye rather than doping in alternating thin layers as was done here. However, a mixture may possess increased losses from Dexter transfer from the sensitizer to the triplet state of the fluorescent dye. To reduce this loss, an ideal system may incorporate low concentrations of a sterically hindered fluorescent dye. For example, adding spacer groups to the DCM2 molecule could decrease the probability of Dexter transfer to the dye while minimally affecting its participation in Förster transfer or its luminescence efficiency. As Dexter transfer can be understood as the simultaneous transfer of an electron and a hole, steric hindrance may also reduce the likelihood of charge trapping on the fluorescent dye. Similar efforts have already reduced non-radiative excimer formation in a DCM2 variant¹⁵. □

Received 19 July; accepted 20 December 1999.

- Baldo, M. A. *et al.* Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature* **395**, 151–154 (1998).
- Baldo, M. A., Lamansky, S., Burrows, P. E., Thompson, M. E. & Forrest, S. R. Very high-efficiency green organic light-emitting devices based on electrophosphorescence. *Appl. Phys. Lett.* **75**, 4–6 (1999).
- VanSlyke, S. A. & Tang, C. W. Organic electroluminescent devices having improved power conversion efficiencies. (US Patent No. 4539507, 1985).
- Baldo, M. A., O'Brien, D. F., Thompson, M. E. & Forrest, S. R. The excitonic singlet-triplet ratio in a semiconducting organic thin film. *Phys. Rev. B* **60**, 14422–14428 (1999).
- Klessinger, M. & Michl, J. *Excited States and Photochemistry of Organic Molecules* (VCH Publishers, New York, 1995).
- Cleave, V., Yahioglu, G., Le Barny, P., Friend, R. & Tessler, N. Harvesting singlet and triplet energy in polymer LEDs. *Adv. Mater.* **11**, 285–288 (1999).
- Chen, C. H., Shi, J. & Tang, C. W. Recent developments in molecular organic electroluminescent materials. *Macromol. Symp.* **125**, 1–48 (1997).
- Brackmann, U. *Lamdachrome Laser Dyes* (Lambda Physik, Göttingen, 1997).
- Forster, T. Transfer mechanisms of electronic excitation. *Discuss. Faraday Soc.* **27**, 7–17 (1959).
- Ermolaev, V. L. & Sveshnikova, E. B. Inductive-resonance transfer of energy from aromatic molecules in the triplet state. *Dokl. Akad. Nauk* **149**, 1295–1298 (1963).
- Berggren, M., Dodabalapur, A., Slusher, R. E. & Bao, Z. Light amplification in organic thin films using cascade energy transfer. *Nature* **389**, 466–469 (1997).
- Garbuzov, D. Z., Bulovic, V., Burrows, P. E. & Forrest, S. R. Photoluminescence efficiency and absorption of aluminium-tris-quinolate (Alq₃) thin films. *Chem. Phys. Lett.* **249**, 433–437 (1996).
- Vander Donckt, E., Camerman, B., Hendrick, F., Herne, R. & Vandeloise, R. Polystyrene immobilized Ir(III) complex as a new material for oxygen sensing. *Bull. Soc. Chim. Belg.* **103**, 207–211 (1994).
- Bulovic, V. *et al.* Bright, saturated, red-to-yellow organic light-emitting devices based on polarization-induced spectral shifts. *Chem. Phys. Lett.* **287**, 455–460 (1998).
- Chen, C. H., Tang, C. W., Shi, J. & Klubeck, K. P. Improved red dopants for organic electroluminescent devices. *Macromol. Symp.* **125**, 49–58 (1997).
- Bulovic, V., Deshpande, R., Thompson, M. E. & Forrest, S. R. Tuning the color emission of thin film molecular organic light emitting devices by the solid state solvation effect. *Chem. Phys. Lett.* **308**, 317 (1999).
- Tang, C. W., VanSlyke, S. A. & Chen, C. H. Electroluminescence of doped organic thin films. *J. Appl. Phys.* **65**, 3610–3616 (1989).
- O'Brien, D. F., Baldo, M. A., Thompson, M. E. & Forrest, S. R. Improved energy transfer in electrophosphorescent devices. *Appl. Phys. Lett.* **74**, 442–444 (1999).

Acknowledgements

This work was supported by Universal Display Corporation, the Defense Advanced Research Projects Agency, the Air Force Office of Scientific Research and the National Science Foundation.

Correspondence and requests for materials should be addressed to S.R.F. (e-mail: forrest@ee.princeton.edu).

Explanation for fracture spacing in layered materials

T. Bai*, D. D. Pollard* & H. Gao†

* Department of Geological and Environmental Sciences,

† Department of Mechanical Engineering, Stanford University, Stanford, California 94305-2115, USA

The spacing of opening-mode fractures in layered materials—such as certain sedimentary rocks and laminated engineering materials—is often proportional to the thickness of the fractured layer^{1–4}. Experimental studies of this phenomenon^{1,5} show that the spacing initially decreases as extensional strain increases in the direction perpendicular to the fractures. But at a certain ratio of spacing to layer thickness, no new fractures form and the additional strain is accommodated by further opening of existing fractures: the spacing then simply scales with layer thickness, which is called fracture saturation^{5,6}. This is in marked contrast to existing theories of fracture, such as the stress-transfer theory^{7,8}, which predict that spacing should decrease with increasing strain *ad infinitum*. Recently^{9,10}, two of us (T.B. and D.D.P.) have used a combination of numerical simulations and laboratory experiments to show that, with increasing applied stress, the normal stress acting between such fractures undergoes a transition from tensile to compressive, suggesting a cause for fracture saturation. Here we investigate the full stress distribution between such fractures, from which we derive an intuitive physical model of the process of fracture saturation. Such a model should find wide applicability, from geosciences^{11–13,14} to engineering^{1,2,6,15,16}.

Spacing of opening-mode fractures in layered materials is important to engineers and geoscientists. Mechanical engineers, civil engineers and materials scientists have studied the spacing of fractures in composite materials for structural safety and service life^{1,2,6,15,16}. Geoscientists have investigated fracture density in layered rocks because of its impact on the flow of groundwater and hydrocarbons^{11–13}, and the safety of mines¹⁴. Experimental results using

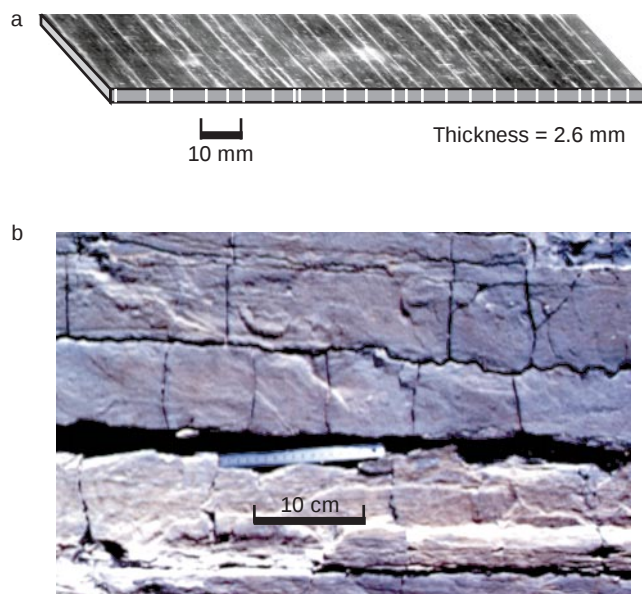


Figure 1 Examples of opening-mode fractures in layered materials. **a**, In a glass-fibre-reinforced polyester (modified from ref. 1). **b**, In the limestone layers of the Carmel Formation, Chimney Rock, Utah.

engineering materials^{1,2,5} (Fig. 1a) and field observations of layered rocks^{3,4,11,13} (Fig. 1b) show that opening-mode fractures often are confined by layer boundaries (Fig. 2) with their height equal to the thickness of the fractured layer, T_f and their spacing, S , proportional to this thickness. Experimental results also show that fracture spacing decreases approximately as the inverse of the applied strain in the direction perpendicular to the fractures, by fractures forming between earlier formed fractures. This process is called sequential infilling⁴. Eventually the fractures are spaced so closely that no more fractures can infill, even with increasing strain. The additional strain is accommodated by further opening of existing fractures. This condition is called fracture saturation^{5,6}.

The commonly used stress-transfer theory⁸ considers a three-layer model with a central layer containing equally-spaced fractures (Fig. 2). Extension parallel to the layers induces a tensile stress, σ_{xx} . Because of the traction-free condition at the fracture surfaces, the stress σ_{xx} in the fractured layer falls to zero at these surfaces. In the region between adjacent fractures, the stress is transferred from the intact neighbouring layers. As the applied strain increases, the stress midway between adjacent fractures will reach the tensile strength; thus new fractures are induced and the spacing is reduced by half. As the strain increases to infinity, the spacing decreases to an infinitesimal value, with no apparent lower limit. Therefore, this theory fails to predict fracture saturation. As shown by the authors⁹, the stress state from the stress-transfer theory does not satisfy the fundamental equations of equilibrium for an elastic boundary-value problem.

The stress-shadow concept¹⁷ is based on the fact that σ_{xx} increases with increasing distance from a single fracture and eventually reaches the remotely applied value. Within a certain distance from the fracture, new fractures cannot form because the stress level is below the tensile strength. This provides a minimum spacing for a given loading. Conceptually, this is appropriate, but the stress level at which new fractures form is usually set arbitrarily, precluding specific applications.

The energy-balance theory^{16,18–20} considers the energy needed for fracture propagation by comparing the strain energy of a portion of the system or the whole system before and after fracturing. The spacing of fractures is limited by the energy available for their formation. This approach usually gives the lower limit of fracture spacing. The results^{19,20} show that spacing decreases nonlinearly with increasing applied strain. As the strain increases to infinity, the

spacing decreases to an infinitesimal value, so this theory fails to predict the experimental results.

There are other theories and conceptual models, which we have reviewed⁹. All of these either give the wrong stress distribution or fail to predict the experimental results from engineering materials and field data on rocks. Therefore, a new theoretical understanding is needed.

Using the three-layer model in Fig. 2, we have investigated the stress distribution between two adjacent opening-mode fractures as a function of the fracture-spacing-to-layer-thickness ratio based on the complete set of governing equations for the elastic boundary-value problem²¹. Four fractures are introduced in the fractured central layer, which is bonded to the adjacent layers. The two fractures in the middle are used to represent any two adjacent fractures in a row composed of many members. Approximating many fractures with only four has been shown to produce errors in calculations of fracture aperture and stress distribution of less than 2% (ref. 9). The average strain in the x -direction is calculated as $\epsilon_{xx}(\text{ave}) = 2u_0/w$, where u_0 is the magnitude of the applied displacements and w is the length of the layers. The middle points of the top and bottom boundaries are fixed in the x -direction, and the middle points of the left and right boundaries are fixed in the y -direction. The top and bottom boundaries are postulated to be traction-free. Also, a plane strain condition for the entire model is postulated, and the elastic moduli of all three layers are identical. We have shown that contrasting moduli introduce minor quantitative differences, but do not alter the qualitative conclusions⁹. We use a two-dimensional finite element code named FRANC (FRacture ANalysis Code) to solve the problem²².

The distributions of the stress σ_{xx} between the two middle fractures (area ABCD, Fig. 2) for different spacing-to-layer-thickness ratios are shown in Fig. 3. The sign convention used here is positive for tension and negative for compression. The stress

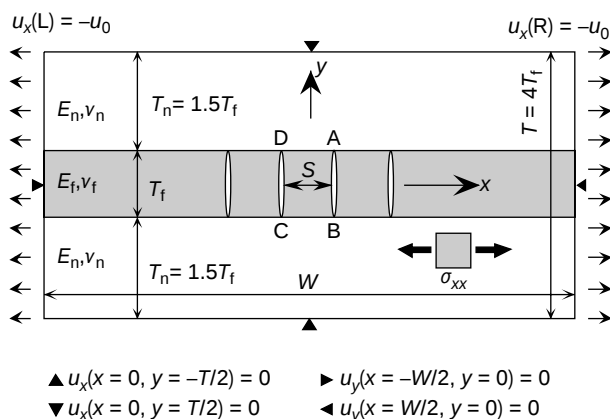


Figure 2 Finite element model and its boundary conditions with four fractures of spacing S in the fractured layer. Displacements are represented by u , and the subscripts x and y indicate the corresponding directions. L , left; R , right. The thicknesses are represented by T . The subscript n indicates the neighbouring layers, and f indicates the fractured layer. E , Young's moduli; ν , Poisson's ratios; σ_{xx} , horizontal stress; W , width of the model.

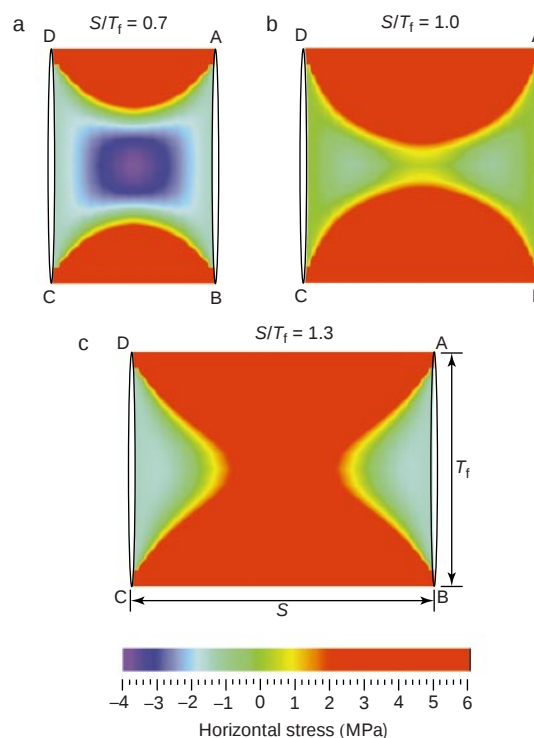


Figure 3 Contour plots of the horizontal stress (σ_{xx}) in the fractured layer between the two middle fractures (area ABCD in Fig. 2) at different fracture-spacing-to-layer-thickness ratios (S/T_f). Note the stress changes sign as the spacing-to-layer-thickness ratios change from less than to greater than a value about 1.0.

distributions are calculated for cases with an average longitudinal strain of $\epsilon_{xx}(\text{ave}) = 0.002$. Figure 3 shows that for fracture-spacing-to-layer-thickness ratios less than about 1.0, there is a region of compressive stress between fractures (Fig. 3a). In contrast, for the ratios greater than 1.0, the stress is tensile throughout this region (Fig. 3c). In other words, the stress state between two fractures changes from tensile to compressive as the ratio changes from greater than to less than a critical value. This value is defined as the critical-spacing-to-layer-thickness ratio⁹.

The stress state transition implies that a new fracture will not fill in between two fractures below the critical value unless a significant flaw locally perturbs the stress field between the fractures, or internal fluid pressure overcomes the compressive stress. In the absence of such mechanisms the stress-state transition provides an explanation for fracture saturation.

The stress distributions shown in Fig. 3 are not intuitive. Given the applied extension and the traction-free fracture surfaces, we would suppose the stress σ_{xx} would decrease toward zero as the fracture-spacing-to-layer-thickness decreased. Indeed, the transition to compressive stress has not been predicted by any of the

known theories and conceptual models. To rebuild an intuitive explanation, we consider that the stress state between two fractures is determined by three mechanisms. The first mechanism is transfer of stress from the neighbouring layers. This can only produce tensile stress in the central area^{7,8} (Fig. 4a, σ_{xx}^T , hollow arrows). The second mechanism involves contraction of the fractured layer in the vertical direction because the fractures shorten as they open. For the same layered system, with no fracture, a longitudinal extension causes a vertical contraction proportional to Poisson's ratio: $\epsilon_{yy} = -\nu\epsilon_{xx}$. In this case, the vertical stress is zero from Hooke's law²¹. However, if fractures are present as shown in Fig. 4a, the fractured layer contracts more than implied by this strain relation. The additional contraction causes a vertical compressive stress, which decreases away from the fracture. Because of the constraint of the surrounding material in the horizontal direction, the vertical compressive stress produces a horizontal compressive stress (Fig. 4a, σ_{xx}^C , solid horizontal arrows). The third mechanism is the effect of the traction-free fracture surface that eliminates both the transferred stress and the effect of horizontal constraint near the fracture.

The horizontal stress between two fractures is determined by the relative magnitudes of the transferred tensile stress, and the compressive stress from fracture shortening and the horizontal constraint, that is, $\sigma_{xx} = \sigma_{xx}^T + \sigma_{xx}^C$. The total stress is controlled by the spacing-to-layer-thickness ratio. When the fracture-spacing-to-layer-thickness ratio is greater than the critical value, the total stress is tensile (Fig. 4b) because more stress can be transferred to the fractured layer, and there is less vertical compression far from the fracture. If the fracture-spacing-to-layer-thickness ratio is less than the critical value, the total stress is compressive (Fig. 4b) because less stress can be transferred and the fractured layer will contract more. To produce the compressive stress, a horizontal constraint must exist. This constraint decreases with decreasing spacing-to-layer-thickness ratio because of the traction-free fracture surface. When the ratio is small enough (<0.3 , Fig. 4b), the horizontal constraint becomes weak. Although there is high compressive stress in the vertical direction, the horizontal compressive stress is small. Furthermore, little stress can be transferred into the narrow region between fractures so the total stress is close to zero.

The stress-transfer theory alone fails to predict the stress-state transition. In fact, the stress state between adjacent fractures predicted by the stress-transfer theory is always tensile^{7,8}. Conceptually, this is because the stress-transfer theory only considers the stress transferred from the neighbouring layers; it does not consider the compressive stress caused by the vertical shortening of the fractures or the horizontal constraint in the central area between two fractures.

We suggest that the stress-transfer theory should be abandoned in future studies of fracture spacing in layered materials. To completely understand fracture spacing, full solutions that are based on the complete set of governing equations for the elastic boundary-value problems should be used. These solutions provide new insights that explain fracture saturation and the quantitative relationships necessary to improve structural design. □

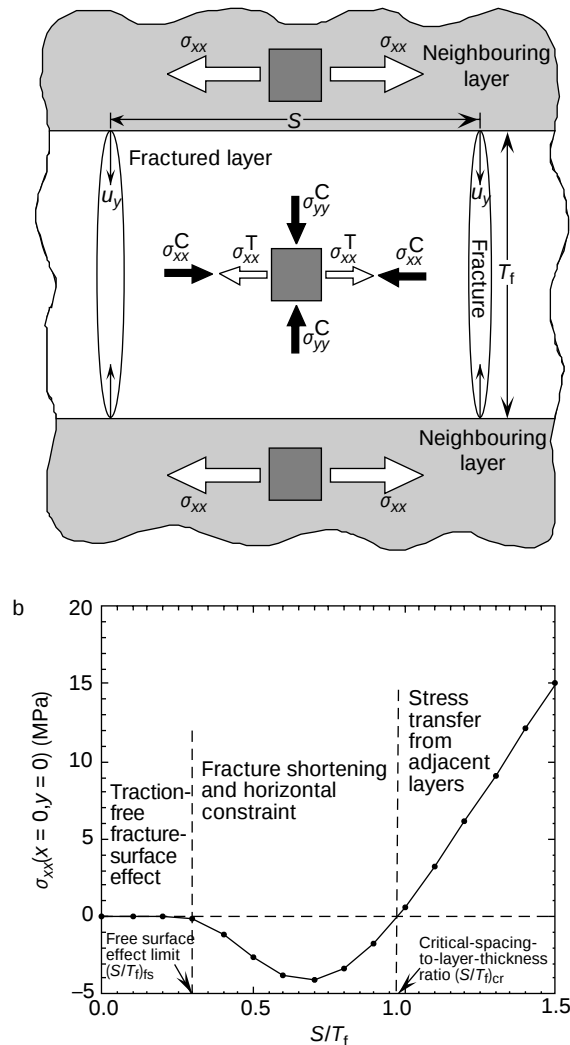


Figure 4 Mechanisms for the stress state transition between adjacent equally-spaced fractures. **a**, Conceptual model of stress state between adjacent fractures. $\sigma_{xx} = \sigma_{xx}^T + \sigma_{xx}^C$, $\sigma_{xx} \geq 0$, if $\sigma_{xx}^T \geq |\sigma_{xx}^C|$; and $\sigma_{xx} < 0$, if $\sigma_{xx}^T < |\sigma_{xx}^C|$. T, tensile; C, compressive. **b**, Three mechanisms that control the horizontal stress (σ_{xx}) at the central point between adjacent fractures as a function of fracture-spacing-to-layer-thickness ratio (S/T_f). When $S/T_f > (S/T_f)_{cr}$, $\sigma_{xx} > 0$; whereas $S/T_f < (S/T_f)_{cr}$, $\sigma_{xx} < 0$. We note that for $S/T_f \leq 0.3$, $\sigma_{xx} \approx 0$. $E_n = E_f = 30,000$ MPa. $\nu_n = \nu_f = 0.25$.

Received 6 July; accepted 21 December 1999.

- Garrett, K. W. & Bailey, J. E. Multiple transverse fracture in 90° cross-ply laminates of a glass fibre-reinforced polyester. *J. Mater. Sci.* **12**, 157–168 (1977).
- Parvizi, A. & Bailey, J. E. On multiple transverse cracking in glass fibre epoxy cross-ply laminate. *J. Mater. Sci.* **13**, 2131–2136 (1978).
- Narr, N. & Suppe, J. Joint spacing in sedimentary rocks. *J. Struct. Geol.* **13**, 1037–1048 (1991).
- Gross, M. R. The origin and spacing of cross joints: examples from Monterey Formation, Santa Barbara Coastline, California. *J. Struct. Geol.* **15**, 737–751 (1993).
- Wu, H. & Pollard, D. D. An experimental study of the relationship between joint spacing and layer thickness. *J. Struct. Geol.* **17**, 887–905 (1995).
- Aveston, J., Cooper, G. A. & Kelly, A. *The Properties of Fiber Composites* 15 (IPC Sci. Technol. Press, London, 1971).
- Cox, H. L. The elasticity and strength of paper and other fibrous materials. *Br. J. Appl. Phys.* **3**, 72–79 (1952).
- Hobbs, D. W. The formation of tension joints in sedimentary rocks: an explanation. *Geol. Mag.* **104**, 550–556 (1967).

9. Bai, T. & Pollard, D. D. Fracture spacing in layered rocks: a new explanation based on the stress transition. *J. Struct. Geol.* **22**, 43–57 (2000).
10. Bai, T. & Pollard, D. D. Spacing of fractures in a multilayer at fracture saturation. *Int. J. Fract.* (in the press).
11. Price, N. J. *Fault and Joint Development in Brittle and Semi-Brittle Rocks* 176 (Pergamon, Oxford, 1966).
12. Pollard, D. D. & Aydin, A. Progress in understanding jointing over the past century. *Geol. Soc. Am. Bull.* **100**, 1181–1204 (1988).
13. National Research Council *Rock Fractures and Fluid Flow: Contemporary Understanding and Applications* 551 (National Academy Press, Washington DC, 1996).
14. Whittaker, B. N., Gaskell, P. & Reddish, D. J. Subsurface ground strain and fracture development associated with longwall mining. *Mining Sci. Technol.* **10**, 71–80 (1990).
15. Thouless, M. D., Olsson, E. & Gupta, A. Cracking of brittle films on elastic substrates. *Acta Metall. Mater.* **40**, 1287–1292 (1992).
16. Hong, A. P., Li, Y. N. & Bazant, P. Theory of crack spacing in concrete pavements. *J. Eng. Mech.* **123**, 267–275 (1997).
17. Lachenbruch, A. H. Depth and spacing of tension cracks. *J. Geophys. Res.* **66**, 4273–4292 (1961).
18. Hu, M. S. & Evans, A. G. The cracking and decohesion of thin films on ductile substrates. *Acta Metall.* **37**, 917–925 (1989).
19. Thouless, M. D. Some mechanics for the adhesion of thin films. *Thin Solid Films* **181**, 397–406 (1989).
20. Thouless, M. D. Crack spacing in brittle films on elastic substrates. *J. Am. Ceram. Soc.* **73**, 2144–2146 (1990).
21. Timoshenko, S. P. & Goodier, J. N. *Theory of Elasticity* 3rd edn, 567 (McGraw-Hill, New York, 1970).
22. Wawrzynek, P. A. & Ingraffea, A. R. Interactive finite element analysis of fracture processes: An integrated approach. *Theor. Appl. Fract. Mech.* **8**, 137–150 (1987).

Acknowledgements

We thank A. Aydin, M. Gross, G. Mavko and Y. Yue for discussions and suggestions. This work was supported by the Stanford RFP and the US NSF.

Correspondence and requests for materials should be addressed to T. B. (e-mail: bai@pangea.stanford.edu).

Temperature trends over the past five centuries reconstructed from borehole temperatures

Shaopeng Huang^{*}, Henry N. Pollack^{*} & Po-Yu Shen[†]

^{*} Department of Geological Sciences, The University of Michigan, Ann Arbor, Michigan 48109-1063, USA

[†] Department of Earth Sciences, The University of Western Ontario, London, Ontario, N6A 5B7, Canada

For an accurate assessment of the relative roles of natural variability and anthropogenic influence in the Earth's climate, reconstructions of past temperatures from the pre-industrial as well as the industrial period are essential. But instrumental records are typically available for no more than the past 150 years. Therefore reconstructions of pre-industrial climate rely principally on traditional climate proxy records^{1–5}, each with particular strengths and limitations in representing climatic variability. Subsurface temperatures comprise an independent archive of past surface temperature changes that is complementary to both the instrumental record and the climate proxies. Here we use present-day temperatures in 616 boreholes from all continents except Antarctica to reconstruct century-long trends in temperatures over the past 500 years at global, hemispheric and continental scales. The results confirm the unusual warming of the twentieth century revealed by the instrumental record⁶, but suggest that the cumulative change over the past five centuries amounts to about 1 K, exceeding recent estimates from conventional climate proxies^{2–5}. The strength of temperature reconstructions from boreholes lies in the detection of long-term trends, complementary to conventional climate proxies, but to obtain a complete picture of past warming, the differences between the approaches need to be investigated in detail.

The thermal regime of the uppermost continental crust is determined in part by the outward flow of heat from the deep interior of the Earth and in part by fluctuations of temperature at the surface. In homogeneous rock and in the absence of changes at the surface, the temperature in the subsurface increases linearly with depth, at a rate which is governed by the magnitude of the terrestrial heat flow and the thermal conductivity of the rock. Fluctuations of surface temperature propagate downward into the rock as attenuating thermal waves superimposed on the temperature profile associated with the deeper heat flow. The depth to which disturbances can be observed is determined by the amplitude, duration and spectral composition of the temperature change at the surface. Owing to the generally low thermal diffusivity of rock, propagation of climate signals in the subsurface is slow. Following a change in temperature at the surface, it takes about 100 years for the perturbation to reach a depth of 150 m, and 1,000 years to reach 500 m depth. Complications in reconstructing a ground surface temperature (GST) history from subsurface temperature data can, however, arise from various non-climatic disturbances⁷ that perturb subsurface temperatures.

We have assembled a database of borehole temperatures for climate reconstruction⁸. The database currently contains 616 borehole temperature profiles that meet certain quality-control criteria; 453 are in the Northern Hemisphere and 163 in the Southern Hemisphere. These borehole sites (Fig. 1) sample all continents except Antarctica, although the geographical distribution of the sites is uneven. The borehole temperatures in this global database were typically measured at 10-m depth intervals to depths as great as 600 m.

The reconstruction of a GST history by inversion of subsurface temperatures has its foundation in the theory of heat conduction^{9–11}. Because of the diffusion of the climate signal through the rocks, a geothermal climate reconstruction is characterized by a progressive inability to resolve the details of climate excursions in the more remote past^{12–14}. For inverting subsurface temperatures to yield a GST history, we use a bayesian estimation technique¹⁵ that is a simplification and extension of the functional space inversion formulation of Shen and Beck¹⁶. Rather than treating the GST history as an arbitrary function of time, we have chosen to parametrize the reconstruction simply, in terms of century-long rates of change over the past five centuries. This simple parametrization leads to a very smooth GST reconstruction. The absence of shorter-period representation in the reconstruction, however, is offset by a reduction of variance in the estimated century-long rates, and by the ease with which these rates may be compared to corresponding quantities that emerge from analyses of the instrumental record. If

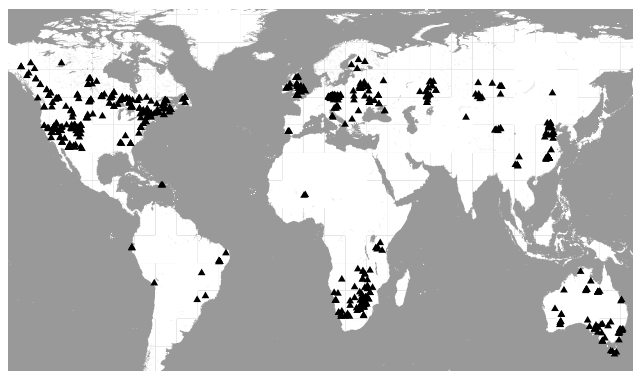


Figure 1 Location map of the boreholes where subsurface temperature measurements have been analysed to reconstruct a ground surface temperature (GST) history. The numbers of boreholes on each continent are respectively 245 (North America), 16 (South America), 146 (Europe), 92 (Africa), 60 (Asia), and 57 (Australia).

climate-change trends do not coincide closely with the calendar centuries, such as when a natural century-long trend straddles two calendar centuries, the inversion will attribute part to one century and part to the other, thus creating a temporal smearing of the temperature trend. In the inversion we employ an *a priori* null hypothesis for the GST history; that is, an initial estimate that there has been no climate change. This is a conservative hypothesis that is also fully independent of any extant models of climate change. As resolution diminishes further back in time, the null hypothesis becomes more difficult to reject.

Of the 616 borehole temperature profiles we analysed, 479 show a net warming over the past five centuries. The average of the cumulative temperature change over the five-century interval is a warming of about 1.0 K (Fig. 2). In the twentieth century alone, the average surface temperature of the continents has increased by about 0.5 K, and the twentieth century has been the warmest century of the past five. This ensemble average is consistent with that derived earlier from a smaller and geographically more restricted data set of 358 boreholes from eastern North America, central Europe, southern Africa, and Australia¹⁷. Although the mechanism of the coupling between the air temperature at the surface and the GST is not simple, and varies from one geographical setting to another^{18–20}, at a large spatial scale the trends of the surface air temperature anomaly and the GST anomaly match well (Fig. 2). Both the global mean surface air temperature (SAT) anomaly series⁶ and the GST continental reconstruction show substantial warming in the twentieth century. The geothermal reconstruction is in generally good agreement with the trend of the global SAT record in both the late nineteenth and twentieth centuries, and extends the climate history back several hundred years before the instrumental

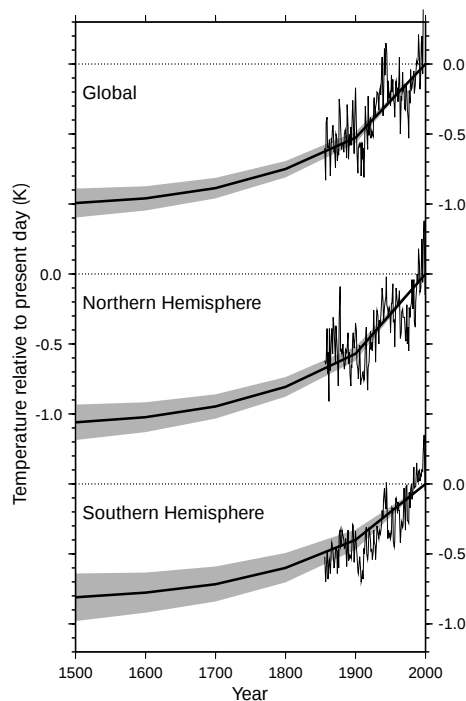


Figure 2 Global and hemispheric averages of GST history over the past five centuries. Shaded areas represent 61 standard error about the mean. Superimposed are the corresponding series of instrumental surface air temperatures (SAT)⁶. Because the geothermal reconstruction is the concatenation of century-long trends, and the SAT anomaly series are referenced to the mean over the period 1961–90, we have shifted the SAT series along the temperature axis to enable an easy comparison of their respective trends. The SAT records have been shifted –0.20 K for the global series, –0.28 K for the Northern Hemisphere and –0.13 K for the Southern Hemisphere.

record. Almost 80% of the net temperature increase observed has occurred in the nineteenth and twentieth centuries.

The magnitude of ground surface warming over the past five centuries is greater in the Northern Hemisphere than in the Southern Hemisphere: the five-century cumulative change is 1.1 K in the former, and 0.8 K in the latter. The twentieth-century temperature change is 0.6 K in the Northern Hemisphere compared with 0.4 K in the Southern Hemisphere. These values compare, respectively, with 0.60 and 0.65 K per century for hemispheric trends in the combined land and sea surface air temperature⁶. The geothermal hemispheric estimates for the twentieth century show even greater consistency with the land-only hemispheric trends of 0.56 and 0.47 K per century reported by Jones²¹. We note that the relatively small number of geothermal observations—and the limited geographical regions represented by them, particularly in the Southern Hemisphere—make tentative any comparisons with hemispheric SAT trends.

Additional regional and temporal variability can be seen in the individual continental GST histories (Fig. 3), although the absence of data in large parts of several continents precludes a detailed interpretation. The five-century cumulative temperature changes are respectively 1.2 K for North America, 1.4 K for South America, 0.8 K for Europe, 0.8 K for Africa, 1.2 K for Asia and 0.5 K for Australia. The GST reconstructions for all six continents exhibit a common characteristic: the temperature change in the twentieth century is the largest of the past five centuries. Intercontinental comparisons for the earlier centuries must be assessed with caution, particularly because of the sparseness of observations in South America and Asia.

Considerable effort has been given recently to combining several proxies in order to produce global, hemispheric and regional-scale climate reconstructions^{1–5,21–23}. Figure 4 shows the comparison of our Northern Hemisphere geothermal reconstruction with three recent multi-proxy representations. All show significant increases of temperature in the twentieth century, but display differences in the previous four centuries. The differences between the various reconstructions may arise in part because of the different geographical distribution of the data used in the respective reconstructions, or perhaps because of differing weights given to individual proxy data sets in the multi-proxy reconstructions²². Of the full hemispheric reconstructions, the geothermal estimate of the five-century temperature change is the largest.

The differences in the pre-instrumental centuries between the geothermal reconstruction and the multi-proxy reconstructions

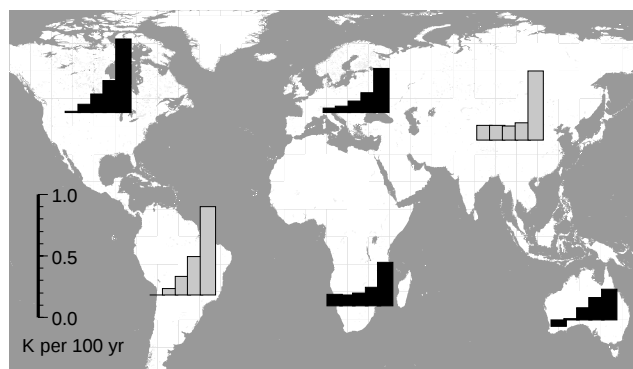


Figure 3 Continental century-long GST changes. In each histogram, the five columns from left to right represent respectively the sixteenth, seventeenth, eighteenth, nineteenth and twentieth centuries. The magnitude of the temperature change is shown as the height of the column. The continental reconstructions for South America and Asia are lightly shaded to indicate the larger uncertainties in these two continents because of the low spatial density of observations.

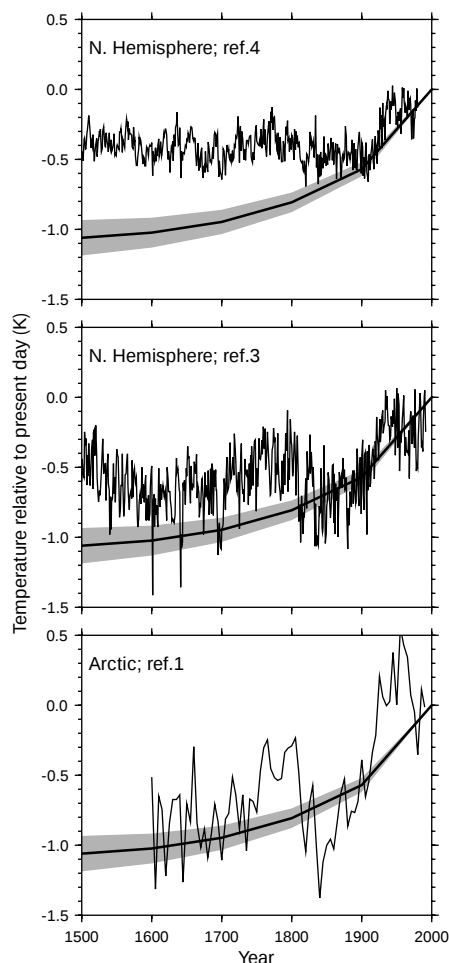


Figure 4 Comparison of five-century Northern Hemisphere geothermal reconstructions with three multi-proxy reconstructions (refs 4, 3 and 1). The Mann *et al.*⁴ and Jones *et al.*³ reconstructions have been shifted along the temperature axis -0.25 K and -0.20 K, respectively, to enable direct comparison of the trends. The Overpeck *et al.*¹ reconstruction has not been shifted.

may also arise in part from the role of tree-ring series in their reconstructions²². Tree-ring data are an important resource in palaeoclimate reconstruction because of their annual resolution and relatively good spatial and temporal coverage. However, tree-ring analyses generally involve some temporal detrending²³, a process that is intended to mute long-term growth trends that may be present in the data. For this reason, the long-term trends derived from borehole temperatures may have a role as useful complements to the traditional proxy reconstructions. Whatever the underlying causes of the differences between the various reconstructions may be, however, the resolution of these differences, particularly in determining the total temperature change over the five-century interval, is important. This temperature change has the potential to be a useful empirical constraint on the climate-sensitivity factor of global climate models. M

Received 7 June; accepted 8 December 1999.

- Overpeck, J. *et al.* Arctic environmental change of the last four centuries. *Science* **278**, 1251–1256 (1997).
- Briffa, K. R., Jones, P. D., Schweingruber, F. H. & Osborn, T. J. Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years. *Nature* **393**, 450–455 (1998).
- Jones, P. D., Briffa, K. R., Barnett, T. P. & Tett, S. F. B. High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with general circulation model control-run temperatures. *Holocene* **8**, 455–71 (1998).
- Mann, M. E., Bradley, R. S. & Hughes, M. K. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* **392**, 779–787 (1998).

- Mann, M. E., Bradley, R. S. & Hughes, M. K. Northern hemisphere temperatures during the past millennium: inferences, uncertainties, and limitations. *Geophys. Res. Lett.* **26**, 759–762 (1999).
- Jones, P. D., New, M., Parker, D. E., Martin, S. & Rigor, I. G. Surface air temperature and its changes over the past 150 years. *Rev. Geophys.* **37**, 173–199 (1999).
- Shen, P.-Y., Pollack, H. N., Huang, S. & Wang, K. Effects of subsurface heterogeneity on the inference of climate change from borehole temperature data: model studies and field examples from Canada. *J. Geophys. Res.* **100**, 6383–6396 (1995).
- Huang, S. & Pollack, H. N. *Global Borehole Temperature Database for Climate Reconstruction* (IGBP PAGES/World Data Center-A for Paleoclimatology Data Contribution Series No. 1998-044, NOAA/NGDC Paleoclimatology Program, Boulder, Colorado, 1998).
- Birch, F. The effects of Pleistocene climatic variations upon geothermal gradients. *Am. J. Sci.* **246**, 729–760 (1948).
- Cermak, V. Underground temperature and inferred climatic temperature of the past millennium. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **10**, 1–19 (1971).
- Lachenbruch, A. H. & Marshall, B. V. Changing climate: geothermal evidence from permafrost in the Alaskan Arctic. *Science* **234**, 689–696 (1986).
- Clow, G. D. Temporal resolution of surface temperature histories inferred from borehole temperature data. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **98**, 81–86 (1992).
- Beltrami, H. & Mareschal, J. C. Resolution of ground temperature histories inverted from borehole temperature data. *Glob. Planet. Change* **11**, 57–70 (1995).
- Harris, R. H. & Chapman, D. S. Climate change on the Colorado Plateau of eastern Utah inferred from borehole temperatures. *J. Geophys. Res.* **103**, 7363–7381 (1998).
- Huang, S., Shen, P.-Y. & Pollack, H. N. Deriving century-long trends of surface temperature from borehole temperatures. *Geophys. Res. Lett.* **23**, 257–260 (1996).
- Shen, P.-Y. & Beck, A. E. Least squares inversion of borehole temperature measurements in functional space. *J. Geophys. Res.* **96**, 19965–19979 (1991).
- Pollack, H. N., Huang, S. & Shen, P.-Y. Climate change record in subsurface temperatures: a global perspective. *Science* **282**, 279–281 (1998).
- Chapman, D. S., Chisolm, T. J. & Harris, R. N. Combining borehole temperature and meteorological data to constrain past climate change. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **98**, 269–281 (1992).
- Lewis, T. J. The effect of deforestation on ground surface temperatures. *Glob. Planet. Change* **18**, 1–13 (1998).
- Majorowicz, J. A. & Skinner, W. R. Potential causes of differences between ground and surface air temperature warming across different ecozones in Alberta, Canada. *Glob. Planet. Change* **15**, 79–91 (1997).
- Jones, P. D. Hemispheric surface air temperature variations: A reanalysis and an update to 1993. *J. Clim.* **7**, 1794–1802 (1994).
- Jones, P. D. It was the best of times, it was the worst of times. *Science* **280**, 544–545 (1998).
- Briffa, K. R. & Osborn, T. J. Seeing the wood from the trees. *Science* **284**, 926–927 (1999).

Acknowledgements

We thank the International Heat Flow Commission and international colleagues for making available many of the borehole temperature profiles that we analysed. S.H. is originally from the Institute of Geology and Geophysics, Chinese Academy of Sciences. This work was supported by the US National Science Foundation, the US National Oceanic and Atmospheric Administration, and the International Geological Correlation Project 428.

Correspondence and request for materials should be addressed to H.N.P. (e-mail: hpollack@umich.edu).

Effect of stream channel size on the delivery of nitrogen to the Gulf of Mexico

Richard B. Alexander, Richard A. Smith & Gregory E. Schwarz

U. S. Geological Survey, 413 National Center, Reston, Virginia 20192, USA

An increase in the flux of nitrogen from the Mississippi river during the latter half of the twentieth century has caused eutrophication and chronic seasonal hypoxia in the shallow waters of the Louisiana shelf in the northern Gulf of Mexico^{1–5}. This has led to reductions in species diversity, mortality of benthic communities and stress in fishery resources⁴. There is evidence for a predominantly anthropogenic origin of the increased nitrogen flux^{2,5–7}, but the location of the most significant sources in the Mississippi basin responsible for the delivery of nitrogen to the Gulf of Mexico have not been clearly identified, because the parameters influencing nitrogen-loss rates in rivers are not well known. Here we present an analysis of data from 374 US monitor-

ing stations, including 123 along the six largest tributaries to the Mississippi, that shows a rapid decline in the average first-order rate of nitrogen loss with channel size—from 0.45 day^{-1} in small streams to 0.005 day^{-1} in the Mississippi river. Using stream depth as an explanatory variable, our estimates of nitrogen-loss rates agreed with values from earlier studies. We conclude that the proximity of sources to large streams and rivers is an important determinant of nitrogen delivery to the estuary in the Mississippi basin, and possibly also in other large river basins.

The problem of tracing nitrogen through large watersheds stems from the difficulty of establishing a spatially continuous mass balance between three rate variables: the in-stream flux of nitrogen, the rate of nitrogen supply from atmospheric and terrestrial sources and the rate of removal due to denitrification and storage on the landscape and in stream channels. Much of the controlled study of supply and removal processes has taken place in relatively small watersheds⁶ where landscape and channel conditions are less variable. Few measurements of nitrogen-loss rates are available for the relatively heterogeneous basins typical of large river channels. Moreover, the reported range of nitrogen-loss rates for stream and river channels exceeds two orders of magnitude, and few explanations for this large variability have emerged. Although various physical and chemical properties of rivers^{6,8–13} are known to influence nitrogen-loss rates, including oxygen concentrations, the organic content of benthic sediments, channel depth, water travel time (that is, water residence time) and stream flow, little has been reported about how loss rates vary over a range of river sizes. In the absence of systematic knowledge of nitrogen-loss rates in channels, no accepted method has emerged for predicting nitrogen transport over long channel distances. Thus, recent efforts^{7,14} to identify the location and types of sources in the Mississippi river basin responsible for nitrogen entering coastal waters have met with only limited success.

We used a recently developed mass-balance method¹² (SPARROW—SPATIALLY-Referenced Regression On Watershed attributes) to estimate nitrogen flux through the interior watersheds of the Mississippi basin, refining the technique to quantify more precisely nitrogen transport in large channels such as those in the Mississippi river and its major tributaries. The regression model and the compilation of the spatial watershed data on nitrogen source inputs, physical characteristics of the landscape and attributes of the digital stream network have been previously described¹² (see Supplementary Information). The method correlates observations of stream nitrogen flux (that is, the response variable) with spatially

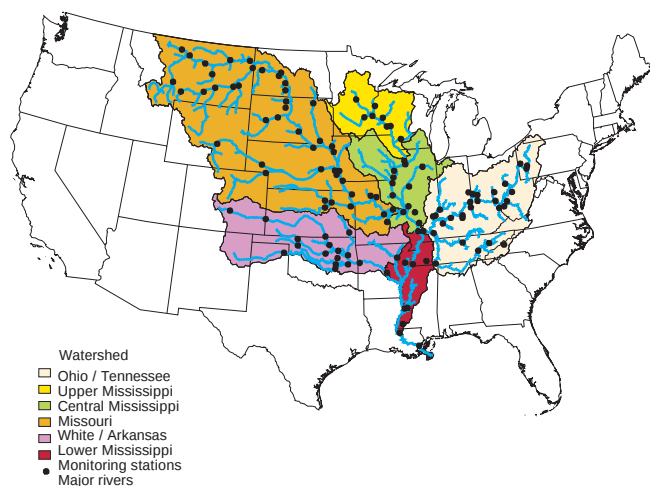


Figure 1 River monitoring stations and major regional watersheds in the Mississippi river basin.

referenced explanatory data on nitrogen source inputs (for example, fertilizer use, atmospheric deposition) and factors controlling nitrogen transport in watersheds, including physical characteristics of the landscape (for example, soil permeability) and aquatic systems (for example, channel size, water velocity). The structural form of the model, which contains separate landscape and stream parameters, provides empirical estimates of the rates of terrestrial and in-stream removal of nitrogen (see Table 1). The response variable in the spatial regression model is mean stream nitrogen flux computed from water-column measurements of total nitrogen (TN; sum of nitrate–nitrite and kjeldahl nitrogen—ammonia plus organic nitrogen) in filtered samples and daily flow measurements¹⁵ at 374 river locations in the United States. These monitoring locations include a subset of 123 stations in the watersheds of six major tributaries to the Mississippi river (Fig. 1). The mean flux estimates at all stations are adjusted to reflect 1987 nitrogen inputs and long-term mean flow conditions, based on the records of concentration and flow for the period 1978 to 1992. The source inputs for 1987 are representative of average inputs over at least the past two decades (see Supplementary Information).

The regression results show that the mean first-order rate of total nitrogen loss (fraction of nitrogen removed per unit water travel

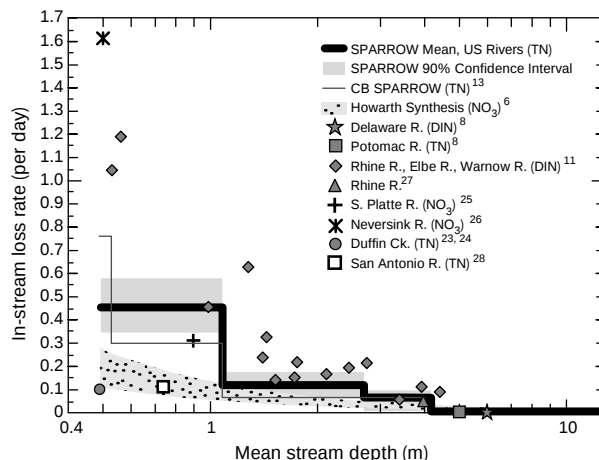


Figure 2 Nitrogen-loss rate in relation to stream channel depth. Stream flow is empirically transformed by regressing observations of mean depth on mean stream flow, on the basis of data from a study of stream morphology and hydraulics at 112 river locations in the United States²⁹ (stream depth and flow are monotonically related according to $D = 0.2612S^{0.3966}$, where D is mean stream depth in metres and S is mean stream flow in $\text{m}^3 \text{ s}^{-1}$; $R^2 = 0.83$). Literature estimates of in-stream nitrogen loss, L (expressed as a fraction of external inputs), are re-expressed as a loss rate (per unit water travel time), R , according to $R = -t^{-1} \ln(1 - L)$, where t is the water time of travel and $\ln$ is the natural logarithm. Loss rates are reported for the following nitrogen forms: total nitrogen (TN), nitrate–nitrogen (NO_3) and dissolved inorganic nitrogen (DIN). SPARROW refers to the mean nitrogen-loss rates estimated in this analysis. CB SPARROW refers to mean estimates separately derived for streams of the Chesapeake Bay watershed over depths of 0.5 to 4 m (ref. 13). Water travel times were estimated for the Rhine, Elbe and Warnow¹¹ rivers from river network data¹² for watersheds in the northeastern quadrant of the United States, based on a regression of the mean water travel time, T , from headwater reaches to the outlet reach of each watershed in units of days on total drainage area, A , in units of square kilometres ($R^2 = 0.88$, $T = -0.0065 + 0.2642 A^{0.3}$). The mean time of travel for watershed streams is computed as one half of T . These estimates assume similar hydraulic properties and drainage density for the European and US watersheds. Channel depths for data from the Rhine river²⁷ and the European watersheds¹¹ were determined from the empirical transformation of stream flow given above. The graphed range of loss rates for nitrate, based on a synthesis of denitrification measurements from selected watersheds in North America⁶, is computed as the ratio of mass transfer coefficients (representing the height of the water column from which nitrate is removed per unit time) to mean depth.

time) is inversely related to channel size, and spans nearly two orders of magnitude (Table 1) over the range of rivers studied. To reliably estimate the functional form over the wide range of river sizes, channel size is defined according to four discrete stream flow classes. Benthic denitrification, a biologically mediated process, is expected to be the dominant loss process quantified by the empirical in-stream loss rates, but the long-term physical storage and release of particulate nitrogen on flood plains and in reservoirs^{16,17} may also be reflected by the rate coefficients. The inverse relation between in-stream nitrogen loss and channel size is probably explained by the influence of channel depth (that is, water stage) on particulate nitrogen settling times and the supply of nitrate for denitrification^{10,18}. The supply of nitrate to benthic sediments is controlled by the direct diffusion of nitrate nitrogen in the water column^{6,8,10,18}, the nitrification of ammonia supplied by mineralized organic nitrogen^{8,19} and the exchange of nitrogen-enriched stream water in the hyporheic zone^{20–22}. Channel depth is a measure of the volume of stream water available for processing by a unit area of benthic sediment. Thus, nitrogen removal by denitrification and settling generally decreases in deeper channels where less contact and exchange of stream waters occurs with the benthic sediment. A previous study⁶ described the percentage aquatic nitrate loss as a declining function of the ratio of depth to water travel time, but did not express nitrogen loss as a rate, and thus did not separate the effect of depth from travel time.

We plotted the empirically derived SPARROW loss rates against mean stream depth in Fig. 2, based on an empirical transformation of stream flow. We compared these estimates of in-stream loss with the available literature estimates^{6,8,11,13,23–28} for the larger temperate watersheds for which estimates of water travel time could be obtained. The travel time of water in streams governs the time of exposure of stream nitrogen to removal processes, including the settling of organic particulate nitrogen, exchange of ground and surface waters (that is, hyporheic flow)^{20–22} and nitrate diffusion to the benthic sediment^{8,18,19}. Few empirical studies of in-stream nitrogen removal are available, and even fewer studies report estimates for total nitrogen and large rivers as analysed in this study. The available literature estimates of percentage nitrogen removal were re-expressed as a loss rate per unit water travel time and plotted as a function of depth. The transformed rates agree reasonably well with the magnitude and declining monotonic pattern of the SPARROW rates of in-stream loss over a wide range

of channel depths (Fig. 2). Similar agreement was also found for in-stream loss rates plotted as a function of stream flow. These findings provide limited confirmation of the validity of the SPARROW rates of nitrogen loss. Moreover, the findings demonstrate the consistency of the literature loss estimates over a considerable range of river sizes when nitrogen loss is expressed as a rate (per unit water travel time) and plotted as a function of depth. In Fig. 2, the greatest agreement in loss rates appears in the larger rivers deeper than one metre, where the literature rates show the least variability and are within a factor of two of the SPARROW rates. For depths above four metres, relatively close agreement is found in the rates of nitrogen loss in the freshwater tidal reaches of the Potomac and Delaware⁸ rivers (0.003 and 0.006 per day, respectively) and the SPARROW rate of 0.005 per day. In rivers with mean depths less than one metre, greater variability generally exists among the literature and SPARROW rates. This may be explained by the greater influence of benthic processes (and variations in benthic conditions, such as the organic content and degree of oxidation of sediments) on nitrogen loss in shallow streams.

Figure 2 indicates that the first-order rate of nitrogen removal in streams (that is, k in Table 1) declines rapidly with increasing channel size as in-stream loss processes become progressively less effective with increases in channel depth. At the basin scale, depth and water velocity increase (as water travel time per unit channel length decreases) in a downstream direction²⁹. Each of these changes in stream attributes contributes to a decrease in stream nitrogen loss per unit channel length. In the Mississippi basin, the decrease in nitrogen loss per unit length due to increases in depth is approximately three times greater than the decrease in loss per unit length due to increases in velocity (see Supplementary Information). To assess the total effect of changes in the first-order loss rate and water velocity on nitrogen delivery to the Gulf of Mexico from interior river locations in the Mississippi basin, we applied the rate coefficients in Table 1 to river network data on channel size and water travel time. We quantified the percentage of the nitrogen export from each of the outlets of 742 interior watersheds in the Mississippi basin that is delivered by streams to the Gulf (Fig. 3; see Supplementary Information for the absolute quantities of nitrogen delivered by source). On the basis of this analysis, we find that the region of high nitrogen delivery to the Gulf is dendritic in shape and extends far upstream along the major tributaries in the eastern and central portions of the basin (for example, Ohio, Tennessee, Lower

Table 1 SPARROW spatial regression model coefficients for total nitrogen

Model parameters	Coefficient units	Bootstrap coefficient	Lower 90% CI	Upper 90% CI
Nitrogen source, β				
Point sources	Dimensionless	0.394	0.094	0.639
Fertilizer application	Dimensionless	1.37	0.605	2.34
Livestock waste production	Dimensionless	0.903	0.012	1.97
Atmospheric deposition	Dimensionless	4.78	1.84	8.21
Non-agricultural land	kg ha ⁻¹ yr ⁻¹	18.6	6.18	29.3
Land-to-water loss coefficient, α				
Temperature	°F ⁻¹	0.017	0.009	0.023
Soil permeability	hr cm ⁻¹	0.036	0.024	0.049
Drainage area per stream length	km ⁻¹	0.043	0.017	0.063
In-stream loss rate coefficient, k				
k_1 (Q , 28.3 m ³ s ⁻¹)	day ⁻¹	0.455	0.344	0.579
k_2 (28.3 m ³ s ⁻¹ , Q , 283 m ³ s ⁻¹)	day ⁻¹	0.118	0.063	0.176
k_3 (283 m ³ s ⁻¹ , Q , 850 m ³ s ⁻¹)	day ⁻¹	0.051	0.007	0.092
k_4 (Q , 850 m ³ s ⁻¹)	day ⁻¹	0.005	0.000	0.019

Coefficients are estimated in a spatial nonlinear least-squares regression of stream nitrogen flux at 374 monitoring locations on watershed characteristics, based on a robust bootstrap estimation procedure¹² ($R^2 = 0.881$; $MSE = 0.435$; see Supplementary Information for explanation of the SPARROW model and coefficients). The non-point-source coefficients (β) multiplied by an exponential land-to-water delivery function (that is, $e^{-\alpha z}$, where z is a vector of land-to-water loss factors; for example, temperature) quantify the proportion of available nitrogen mass delivered to rivers as a function of the specified source inputs and landscape characteristics. The land-to-water delivery function is equal to one for point-source inputs. The rate coefficients (k) quantify the first-order rate of in-stream nitrogen loss per unit of water travel time (for example, $k_4 = 0.5\%$ removal of nitrogen per day of water travel time). The regression residuals provide acceptable adherence to model assumptions. In-stream loss rates are fitted separately for stream reaches with mean stream flow (Q) corresponding to the indicated intervals. Nitrogen-loss rates (for example, k_4) were estimated according to a continuous function of stream flow in preliminary analyses; however, a discrete functional form defined by separate stream flow classes provides the most accurate fit to the observational data for the extreme river sizes; the uppermost class has great significance to large rivers in the Mississippi basin. Model predictions of flux are typically within 32% of the observed values based on the median of station values (interquartile range of 15–61%). A validation of the model predictions through comparisons with an independent data set of 68 monitoring stations in the Chesapeake Bay watershed in the eastern United States provided reasonable confirmation that the predictions are relatively unbiased (model predictions are typically within 39% of the observed values with an interquartile range of 19–82%), and that the residuals accurately describe the unexplained variability in the model from which confidence intervals (CI) are developed.

Missouri, Lower Arkansas and Upper Mississippi rivers). Despite the long water travel times, many watersheds located on large rivers more than 2,500 kilometres from the Gulf deliver significantly larger fractions of their exported nitrogen (some more than 90%) to coastal waters than watersheds located on smaller streams less than a few hundred kilometres from the Gulf. In addition, the dendritic pattern of nitrogen transport leads to widely varying delivery percentages in each of the major regional drainages of the Mississippi basin, ranging from more than 90% from watersheds on the largest rivers to substantially less than 40% from watersheds on small streams (see Table 3 in Supplementary Information). This wide variation is evident despite similarities of the distances of interior watersheds from the Gulf of Mexico within each regional drainage. Nitrogen deliveries from many arid watersheds in the more distant drainages of the western Mississippi basin (that is, the western portions of the Missouri and Arkansas/Red regions) are uniformly small because of the effect of the typically shallow rivers with high nitrogen-loss rates and the lengthy water travel times to the Gulf.

We conclude that, because the nitrogen-loss rate in streams declines rapidly with increasing channel size, knowledge of the length of time that surface waters are transported through channels of varying size can help to predict the quantities of nitrogen delivered from interior locations to coastal waters. Despite uncertainty in the rate of nitrogen loss in stream channels of a given size, the evidence of a large, systematic decline in the rate of nitrogen removal from small streams to large rivers has important implications for nutrient management in the Mississippi river basin, and more generally, in large coastal watersheds. The delivery of nitrogen to coastal systems from point and diffuse sources is not a simple function of the distance of these sources from coastal waters. Instead, the proximity of sources to large streams and rivers, as measured by the length of time that surface waters travel through smaller tributaries, is a major determinant of their downstream transport to marine systems. Information on the rates of nitrogen delivery to coastal waters may assist in evaluations of efficient nutrient control strategies, including efforts to identify the most

significant watersheds and sources contributing to riverine exports of nitrogen to coastal ecosystems. **M**

Received 27 July; accepted 22 December 1999.

1. Rabalais, N. N. *et al.* Nutrient changes in the Mississippi River and system responses on the adjacent continental shelf. *Estuaries* **19**, 386–407 (1996).
2. Turner, R. E. & Rabalais, N. N. Coastal eutrophication near the Mississippi river delta. *Nature* **368**, 619–621 (1994).
3. Vitousek, P. M. *et al.* Human alteration of the global nitrogen cycle: sources and consequences. *Ecol. Appl.* **7**, 737–750 (1997).
4. Diaz, R. J. & Rosenberg, R. Marine benthic hypoxia: a review of its ecological effects and the behavioural responses of benthic macrofauna. *Oceanogr. Mar. Biol. (Annu. Rev.)* **33**, 245–303 (1995).
5. Turner, R. E. & Rabalais, N. N. Changes in Mississippi River water quality this century: implications for coastal food webs. *Bioscience* **41**, 140–147 (1991).
6. Howarth, R. W. *et al.* Regional nitrogen budgets and riverine N & P fluxes for the drainages to the North Atlantic Ocean: natural and human influences. *Biogeochemistry* **35**, 75–139 (1996).
7. Goolsby, D. A. *et al.* Flux and sources of nutrients in the Mississippi-Atchafalaya River basin. (Report of Task Group 3 to the White House Committee on Environment and Natural Resources, Hypoxia Work Group) *Federal Register* **64**, 23834–23835 (1999).
8. Seitzinger, S. P. Denitrification in freshwater and coastal marine ecosystems: ecological and geochemical significance. *Limnol. Oceanogr.* **33**, 702–724 (1988).
9. Seitzinger, S. P. & Kroeze, C. Global distribution of nitrous oxide production and N inputs in freshwater and coastal marine ecosystems. *Glob. Biogeochem. Cycles* **12**, 93–113 (1998).
10. Kelly, C. A. *et al.* Prediction of biological acid neutralization in acid-sensitive lakes. *Biogeochemistry* **3**, 129–141 (1987).
11. Behrendt, H. Inventories of point and diffuse sources and estimated nutrient loads—a comparison for different river basins in central Europe. *Wat. Sci. Tech.* **33**, 99–107 (1996).
12. Smith, R. A., Schwarz, G. E. & Alexander, R. B. Regional interpretation of water-quality monitoring data. *Wat. Resour. Res.* **33**, 2781–2798 (1997).
13. Preston, S. D. & Brakebill, J. W. *Application of Spatially Referenced Regression Modeling for the Evaluation of Total Nitrogen Loading in the Chesapeake Bay watershed*. 12 (US Geological Survey Water Resources Investigations Report 99-4054, Baltimore, Maryland, 1999).
14. Brezonik, P. L. *et al.* Effects of reducing nutrient loads to surface waters within the Mississippi River Basin and the Gulf of Mexico. (Report of Task Group 4 to the White House Committee on Environment and Natural Resources, Hypoxia Work Group) *Federal Register* **64**, 23834–23835 (1999).
15. Alexander, R. B., Slack, J. R., Ludtke, A. S., Fitzgerald, K. K. & Schertz, T. L. Data from selected U. S. Geological Survey national stream water quality monitoring networks. *Wat. Resour. Res.* **34**, 2401–2405 (1998).
16. Johnston, C. A., Bubenzer, G. D., Lee, G. B., Madison, F. W. & McHenry, J. R. Nutrient trapping by sediment deposition in a seasonally flooded lakeside wetland. *J. Environ. Qual.* **13**, 283–290 (1984).
17. Billen, G., Dessery, S., Lancelot, C. & Meybeck, M. Seasonal and inter-annual variations of nitrogen diagenesis in the sediments of a recently impounded basin. *Biogeochemistry* **8**, 73–100 (1989).
18. Baker, L. A. & Brezonik, P. L. Dynamic model of internal alkalinity generation: calibration and application to precipitation-dominated lakes. *Wat. Resour. Res.* **24**, 65–74 (1988).
19. Novotny, V. & Olem, H. *Water Quality: Prevention, Identification, And Management Of Diffuse Pollution* (Van Nostrand Reinhold, New York, 1994).
20. McMahon, P. B. & Bohlke, J. K. Denitrification and mixing in a stream-aquifer system: effects on nitrate loading to surface water. *J. Hydrol.* **186**, 105–128 (1996).
21. Triska, F. J., Duff, J. H. & Avanzino, R. J. Patterns of hydrological exchange and nutrient transformation in the hyporheic zone of a gravel-bottom stream: examining terrestrial-aquatic linkages. *Freshwat. Biol.* **29**, 259–274 (1993).
22. Harvey, J. W., Wagner, B. J. & Bencala, K. E. Evaluating the reliability of the stream tracer approach to characterize stream-subsurface water exchange. *Wat. Resour. Res.* **32**, 2441–2451 (1996).
23. Hill, A. Denitrification in the nitrogen budget of a river ecosystem. *Nature* **281**, 291–292 (1979).
24. Hill, A. Nitrate-nitrogen flux and utilization in a stream ecosystem during low summer flows. *Can. Geogr.* **XXV**, 225–239 (1981).
25. Sjödin, A. L., Lewis, W. M. Jr. & Saunders, J. F. III Denitrification as a component of the nitrogen budget for a large plains river. *Biogeochemistry* **39**, 327–342 (1997).
26. Burns, D. A. Retention of NO_3 in an upland stream environment: a mass balance approach. *Biogeochemistry* **40**, 73–96 (1998).
27. Billen, G., Lancelot, C. & Meybeck, M. *N, P, and Si Retention Along Aquatic Continuum From Land To Ocean, In Ocean Margin Processes In Global Change* (eds Mantoura, R. F. C., Martin, J. M. & Wollast, R.) 19–44 (Wiley, 1991).
28. Twidwell, S. R. & Davis, J. R. *Intensive Surveys of San Antonio River Segments 1901 and 1911, June 6, 1984 – May 16, 1985* (Texas Natural Resource Conservation Commission, IS 87–04, Austin, Texas, 1987).
29. Leopold, L. B. & Maddock, T. Jr. *The Hydraulic Geometry of Stream Channels and Some Physiographic Implications*. (U. S. Geological Survey Professional Paper 252, Reston, Virginia, 1953).
30. Seaber, P. R., Kapinos, F. P. & Knapp, G. L. *Hydrologic Units Maps* (U. S. Geological Survey Water Supply Paper 2294, Reston, Virginia, 1987).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. Bohlke, D. Goolsby, R. Hirsch, S. Seitzinger and N. Rabalais for comments on the manuscript. S. Seitzinger and R. Stiles assisted in obtaining published data from watershed studies of nitrogen loss. M. Ierardi assisted in preparing the figures.

Correspondence and requests for materials should be addressed to R. B. A. (e-mail: ralex@usgs.gov).

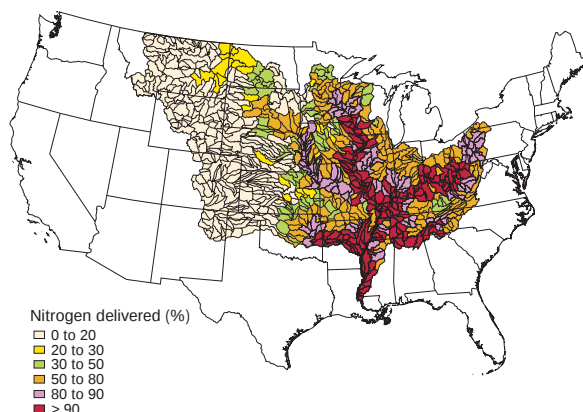


Figure 3 Percentage of the nitrogen export from interior watersheds delivered to the Gulf. Approximately equally sized interior watersheds, ranging from about 2,400 to 4,900 km² (mean is 3,900 km²), are systematically defined according to the hydrologic cataloguing unit classification³⁰. The delivery percentage is the fraction of the nitrogen exported from inland watersheds that remains after in-stream transport to the Gulf, and is computed as $[\exp(-k't)/100]$, where k' is a vector of SPARROW estimates of in-stream nitrogen loss for four stream sizes (Table 1), and t is a vector of mean water travel times from each watershed outlet to the Gulf for each of the four stream sizes. The water travel times from locations above the diversion from the Lower Mississippi river to the Atchafalaya river are computed as the flow-weighted mean of the travel time (2.4 days) for each pathway to the Gulf. See the Supplementary Information for regional estimates of the delivery percentages and the absolute quantities of nitrogen delivered by source.

Producer–decomposer co-dependency influences biodiversity effects

Shahid Naeem, Daniel R. Hahn & Gregor Schuurman

Department of Zoology, University of Washington, 24 Kincaid Hall, Box 351800, Seattle, Washington 98195, USA

Producers, such as plants and algae, acquire nutrients from inorganic sources that are supplied primarily by decomposers whereas decomposers, mostly fungi and bacteria, acquire carbon from organic sources that are supplied primarily by producers. This producer–decomposer co-dependency is important in governing ecosystem processes^{1–4}, which implies that the impacts of declining biodiversity on ecosystem functioning^{5–7} should be strongly influenced by this process. Here we show, by simultaneously manipulating producer (green algal) and decomposer (heterotrophic bacterial) diversity in freshwater microcosms, that algal biomass production varies considerably among microcosms (0.0–0.67 mg ml^{–1}), but that neither algal nor bacterial diversity by itself can explain this variation. Instead, production is a joint function of both algal and bacterial diversity. Furthermore, the range in algal production in microcosms in which bacterial diversity was manipulated was nearly double (1.82 times) that of microcosms in which bacterial diversity was not manipulated. Measures of organic carbon use by bacteria in these microcosms indicate that carbon usage is the mechanism responsible for these results. Because both producer⁸ and microbial diversity⁹ respond to disturbance and habitat modification, the main causes of biodiversity loss⁸, these results suggest that ecosystem response to changing biodiversity is likely to be more complex than other studies^{5–7} have shown.

Eukaryotic, photo-autotrophic producers generally cannot acquire carbon or nutrients from organic sources, but must transform inorganic carbon, such as CO₂, and inorganic nutrients, such as NO₃, into the organic materials needed for growth and reproduction (Fig. 1). Heterotrophic decomposers, in contrast, cannot acquire carbon from inorganic sources, but can transform organic carbon from sources supplied by producers, either as exudates^{4,10–13} or as dead organic matter, and can acquire nutrients from organic or inorganic forms (Fig. 1). Decomposers transform organic nutrients into inorganic forms (mineralization), a main pathway of inorganic nutrients to producers^{2,14} (Fig. 1). The asymmetry in metabolic capabilities results in a mutual co-dependency that inextricably links producers with decomposers and forms the foundation of material cycling in most solar-based ecosystems (Fig. 1). Although these links between autotrophic production, producer biomass, decomposer production and decomposer biomass are well studied

(see for example refs 1, 4, 15, 16), little is known about the role of producer and decomposer diversity in these processes^{17–19}.

To test whether this producer–decomposer co-dependency influences producer diversity impacts on ecosystem functioning, we simultaneously manipulated unicellular green algal (producer) species richness (S_A) and heterotrophic bacterial (decomposer) species richness (S_B) in 116 freshwater, microbial microcosms. All microcosms contained bacteria, but our manipulations supplemented bacterial diversity by adding up to 12 additional species. All microcosms were inoculated weekly with the initial complement to ensure that treatment-designated species were present throughout the experiment. We measured algal production (standing algal biomass), decomposer production (standing microbial biomass) and carbon-source usage as response variables. Carbon source usage was assayed by monitoring the consumption by bacterial communities in each microcosm of 95 different organic carbon sources that span a wide breadth of bacterial metabolic capabilities²⁰. Algae were removed by filtration before the assay, and assays were conducted in the dark to ensure that carbon use was strictly a factor of bacteria present in the microcosm at the time of measurement.

In microcosms where bacteria were not added, algal production showed a positive association with S_A ($R^2 = 0.49$, $P = 0.05$). This result agrees with studies that manipulate only producer diversity (for example, see refs 5, 20).

In contrast, where bacterial species were added, algal production was significantly affected by the interaction between S_A and S_B ($S_A \times S_B$) (Table 1, Fig. 2). Bacterial production did not show a significant response to S_A , but S_B and $S_A \times S_B$ were significant (Table 1, Fig. 2). $S_A \times S_B$ affected total (algal plus bacterial) production (Table 1), although this is probably due to the disproportionate influence of algal biomass on total biomass (Fig. 2). Scaling algal and bacterial biomass to a maximum value of 1, and examining the ratio of these scaled abundances to avoid the disproportionate effect of algal biomass showed a significant $S_A \times S_B$ effect (Table 1).

The number of carbon sources used by bacteria was also affected by S_A and S_B , but not $S_A \times S_B$ (Table 1, Fig. 2). The number of carbon sources used showed a significant positive correlation with S_B ($R = 0.39$, $P = 0.01$), but no significant correlation with either S_A ($R = 0.18$, $P = 0.15$) or total microcosm biomass ($R = 0.15$, $P = 0.51$) was observed. These results indicate that the diversity of carbon sources used in a microcosm is a function of S_B . Although the diversity of carbon sources produced is expected to correlate with S_A , no correlation was expected between S_A and the number of carbon sources used because producers do not use organic carbon (Fig. 1). The significant treatment effects of S_A on the mean number of carbon sources used by bacteria (Table 1), however, indicate that the number of organic carbon sources used is indirectly affected by the impacts of producer diversity on bacterial communities.

The effects of S_A and S_B on numbers of organic carbon sources used by the bacterial communities indicate that the diversity impacts on production may be attributable to carbon-based producer–decomposer interactions that are well known in nature. A significant fraction (5–50%) of the organic carbon produced by

Table 1 Effects of variation in algal species richness, bacterial species richness and the interaction between these factors on microcosm properties

SOURCE	S_A			S_B			$S_A \times S_B$		
	d.f.	F	P	d.f.	F	P	d.f.	F	P
Algal biomass	3	13.6	, 0.001*	4	4.5	, 0.01*	12	5.4	, 0.001*
Bacterial biomass	3	2.5	0.065	4	5.6	, 0.01*	12	3.2	, 0.01*
Total biomass	3	13.3	, 0.001*	4	3.9	, 0.01*	12	5.8	, 0.001*
Ratio of standardized bacterial to algal biomass	3	45.9	, 0.001*	4	10.0	, 0.001*	12	8.6	, 0.001*
No. of carbon sources used	3	7.2	, 0.01*	4	4.6	, 0.01*	12	1.7	0.118

The statistical test used was two-factor analysis of variance (ANOVA). The ratio of standardized bacterial to algal biomass was arcsine-square root transformed. Standardization involved scaling both maximum algal and maximum bacterial biomass to 1 to minimize the influence of algal biomass which was generally larger than bacterial biomass. S_A , algal species richness; S_B , bacterial species richness; $S_A \times S_B$, interaction. F = F value. P = significance. d.f. = degrees of freedom. Degrees of freedom for error term in all analyses was 59.

* Statistical significance.

plant and algae is exuded and consumed by decomposers in nature^{2,4,12}, and the remaining portion is eventually consumed as dead organic matter (although some fraction may become recalcitrant as humic substances or lost because of burial²¹). This organic carbon is a rich variety of complex organic compounds^{3,4,11,15} that is matched by an equally diverse ability of decomposer microbes to consume these compounds²². Our findings are congruent with these observations from nature, thereby supporting bacterial exploitation of organic carbon sources as the link between decomposer diversity and algal production. Stoichiometric relationships suggest that such feedbacks are both carbon and nutrient based^{3,23}, thus the complexity of producer–decomposer feedbacks on diversity–ecosystem relationships may be even stronger than we observed if nutrients vary.

These results support a potentially strong influence of the producer–decomposer co-dependency on ecosystem response to changing biodiversity. Only in the absence of experimental manipulation of decomposer diversity was producer production positively associated with producer diversity; otherwise production was a joint function of producer and decomposer diversity. Producer-only studies (for example, see refs 5, 7, 18) are likely to underestimate the magnitude and complexity of biodiversity impacts. Simulta-

neous variation in decomposer diversity is likely to accompany most changes in diversity that are caused by disturbance or habitat modification^{8,24}. Such linkages as the producer–decomposer co-dependency among species implies that the preservation of biodiversity as a means for preserving, restoring and managing ecosystem functioning requires a multitrophic approach in which decomposers, producers and other trophic groups^{25,26} are integrated in both research and practice.

M

Methods

Microcosms

Microcosms consisted of disposable, sterile 100-mm diameter × 25-mm deep polystyrene Petri dishes each filled with 50 ml of growth medium (Carolina Biological Supply) (for additional detail, see ref. 6). Chamber conditions were 20 °C with a 14:10 h, light:dark photoperiod.

Experimental design

Algal species richness treatments were 0, 1, 2, 4 and 8 species chosen randomly from *Chlamydomonas reinhardtii*, *Carteria olivieri*, *Closterium* sp., *Eremosphaera* sp., *Euglena gracilis*, *Netrium* sp., *Phacus* sp. and *Trachelomonas* sp. Bacterial species richness treatments were 0, 1, 2, 4, 8 and 12 species chosen from a pool consisting of *Aeromonas sobria*, *Aquaspirillum itersonii*, *Aq. serpens*, *Aq. sinuosum*, *Bacillus cereus*, *B. megaterium*, *B.*

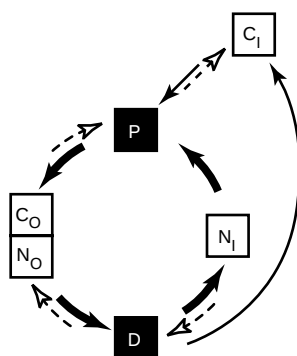


Figure 1 The fundamental producer–decomposer co-dependency common to most ecosystems. Filled boxes represent biomass and open boxes represent material pools. The biomass compartments shown are producers (P) and decomposers (D). The material pools are organic carbon (C_O), inorganic carbon (C_I), organic nutrients (N_O) and inorganic nutrients (N_I). Solid arrows indicate major flows, whereas dashed, open arrows indicate minor flows such as producer resorption of carbon (known to occur in some plants) and

producer respiration. Producers are dependent on decomposers for uptake of inorganic nutrients, whereas decomposers are dependent on producers for organic materials. The cycling of material (counterclockwise array of solid arrows in centre) results from these fundamental activities. (For simplicity, we have treated this system as a system closed to internal and external inputs and outputs.)

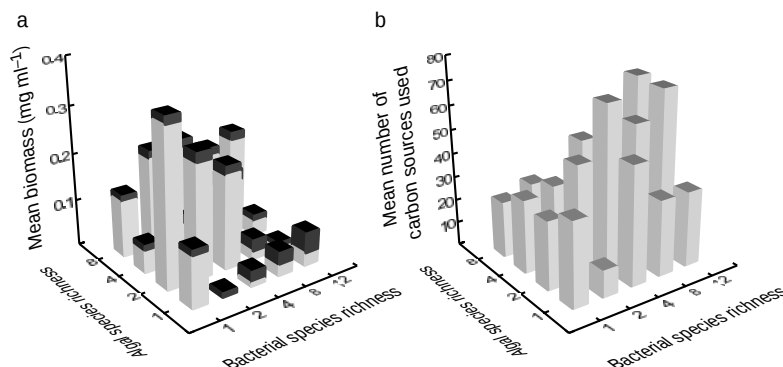


Figure 2 Relationships between producer diversity, bacterial diversity, standing biomass and carbon-source usage. **a**, Mean biomass in microcosm replicates containing different numbers of algal and bacterial species after 4 weeks. Bars are divided with top, darker portion showing the fraction of biomass found in bacterial populations. The highest proportions of bacterial biomass occur when bacterial species richness is high and algal species richness is low, whereas higher proportions of algal biomass occur when algal species richness is intermediate to high and bacterial species richness is low. **b**, Mean

number of carbon sources used in replicate microcosms for each diversity treatment. Carbon sources consisted of 95 unique organic substrates used by bacterial assemblages taken from microcosms. All species of bacteria were freshwater species adapted to culture conditions, thus the problem with unculturables does not apply here. The diversity of carbon sources used is higher when bacterial species richness is high and algal species richness is high to intermediate.

subtilis, *Enterobacter cloacae*, *Ochrobactrum anthropi*, *Pseudomonas aeruginosa*, *Serratia marcescens* and *S. liquefaciens*. We did not construct 0 algae/0 bacteria microcosms. This design yielded 118 microcosms ($S_A = 5$ levels $3 S_B = 6$ levels $3 A$ 4 replicates $2 S_A$ or $S_B = 0$). Algal and bacterial stock cultures were maintained in the same medium used in microcosms. We carried out four replicates per species composition, and replicates were dispersed among four separate growth chambers to avoid pseudoreplication. One hundred algal cells in total were added to each microcosm (for example, 100 cells per species in monocultures, 50 cells per species in two species treatments). One millilitre of bacteria culture(s) was added to each microcosm (for example, 1 ml for each species in monocultures, 0.5 ml of each species in two species treatments). Note that the initial biomass in the small inoculum was insignificant, thus final measures of biomass were used to estimate production.

Measurements

Algal biomass was determined from algal biovolume²⁷, whereas bacterial biomass was determined from conversion of densities determined by DAPI enumeration²⁸ in 5-ml samples. Measurements were made after 4 weeks (algal densities stabilize after two weeks²⁹). Sole source carbon usage was determined by Biolog plate methods¹⁹. Samples were filtered with 20-mm polycarbonate membrane filters to remove algae before plating and plates were incubated at 20 °C for 24 hours. A M3E Biolog plate reader was used to read microplates. The bacterial species that we selected are well adapted to laboratory conditions and pre-tests showed that they grew well in microplates; thus problems associated with unculturable species from environmental samples³⁰ do not apply to our application.

Received 10 September; accepted 18 November 1999.

- Harte, J. & Kinzig, A. P. Mutualism and competition between plants and decomposers: implications for nutrient allocation in ecosystems. *Am. Nat.* **141**, 839–846 (1993).
- Kaye, J. P. & Hart, S. C. Competition for nitrogen between plants and soil microorganisms. *Trends Ecol. Evol.* **12**, 139–143 (1997).
- Berman-Frank, I. & Dubinsky, Z. Balanced growth in aquatic plants: myth or reality? *BioScience* **49**, 29–37 (1999).
- Wall, D. H. & Moore, J. C. Interactions underground. *BioScience* **49**, 109–117 (1999).
- Tilman, D. *et al.* The influence of functional diversity and composition on ecosystem processes. *Science* **277**, 1300–1302 (1997).
- Naeem, S. & Li, S. Consumer species richness and autotrophic biomass. *Ecology* **79**, 2603–2615 (1998).
- Hooper, D. U. & Vitousek, P. M. The effects of plant composition and diversity on ecosystem processes. *Science* **277**, 1302–1305 (1997).
- Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. Quantifying threats to imperiled species in the United States. *BioScience* **48**, 607–615 (1998).
- Borneman, J. & Triplett, E. W. Molecular microbial diversity in soils from Eastern Amazonia: Evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Applied Environ. Microbiol.* **63**, 2647–2653 (1998).
- Griffiths, B. S., Ritz, K. & Wheatley, R. E. in *Microbial Communities* (eds Insam, H. & Rangger, A.) 1–9 (Springer, Berlin, 1997).
- Wood, A. M. & Van Valen, L. M. Paradox lost? On the release of energy-rich compounds by phytoplankton. *Marine Microbial Food Webs* **4**, 103–116 (1990).
- Jones, A. K. The interaction of algae and bacteria. in *Microbial interactions and communities* (eds Bull, A. T. & Slater, J. H.) 189–247 (Academic, London, 1982).
- Berman, T. & Holm-Hanse, O. Release of photoassimilated carbon as dissolved organic matter by marine phytoplankton. *Mar. Biol.* **28**, 305–310 (1974).
- Zak, D. R. & Grigal, D. F. Nitrogen mineralization, nitrification and denitrification in upland and wetland ecosystems. *Oecologia* **88**, 189–196 (1991).
- del Giorgio, P. A. & Cole, J. J. Bacterial growth efficiency in natural aquatic systems. *Annu. Rev. Ecol. System.* **29**, 503–541 (1998).
- Zak, D. R., Grigal, D. F., Gleeson, S. & Tilman, D. Carbon and nitrogen cycling during old-field succession: constraints on plant and microbial biomass. *Biogeochemistry* **11**, 111–129 (1990).
- Hobbie, S. E. Effects of plant species on nutrient cycling. *Trends Ecol. Evol.* **7**, 336–339 (1992).
- Naeem, S., Haakenson, K., Thompson, L. J., Lawton, J. H. & Crawley, M. J. Biodiversity and plant productivity in a model assemblage of plant species. *Oikos* **76**, 259–264 (1996).
- Allsopp, D., Colwell, R. R. & Hawksworth, D. L. (eds) *Microbial Diversity and Ecosystem Function* (CAB International, Oxon, 1995).
- Bochner, B. R. Sleuthing out bacterial identities. *Nature* **339**, 157–158 (1989).
- Schlesinger, W. H. *Biogeochemistry 3rd edn* (Academic, San Diego, 1997).
- Pritchard, P. H., Mueller, J. G., Lantz, S. E. & Santavay, D. L. in *Microbial Diversity and Ecosystem Function* (eds Allsopp, D., Colwell, R. R. & Hawksworth, D. L.) 161–182 (CAB International, Oxon, 1995).
- Elser, J. J., Dobberfuhl, D. R., MacKay, N. A. & Schampel, J. H. Organism size, life history and N:P stoichiometry. *BioScience* **46**, 674–684 (1996).
- Lawton, J. H. *et al.* Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature* **391**, 72–79 (1998).
- Mulder, C. P. H., Koricheva, J., Huss-Danell, K., Höglberg, P. & Joshi, J. Insects affect relationships between plant species richness and ecosystem processes. *Ecol. Lett.* **2**, 237–246 (1999).
- Van der Heijden, M. G. A. *et al.* Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* **396**, 69–72 (1998).
- Wetzel, R. G. & Likens, G. E. *Limnological Analyses* (Springer, Berlin, Germany, 1991).
- Porter, K. G. & Feig, Y. S. The use of DAPI for identifying and counting aquatic microflora. *Limnol. Oceanogr.* **25**, 943–948 (1980).
- Naeem, S. & Li, S. Biodiversity enhances ecosystem reliability. *Nature* **390**, 507–509 (1997).
- Smalla, K., Wachtendorf, U., Heuer, H., Liu, W.-T. & Forney, L. Analysis of BIOLOG GN substrate utilization patterns by microbial communities. *Applied Environ. Microbiol.* **64**, 1220–1225 (1998).

Acknowledgements

We thank B. Bohanan, M. Brett, J. Cotner, A. Grack, S. Lawler, J. Lawton, P. Morin, D. Schindler and S. Tjossem for help, comments and discussion. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to S.N. (e-mail: naeems@u.washington.edu).

The mouse *Dreher* gene *Lmx1a* controls formation of the roof plate in the vertebrate CNS

James H. Millonig^{*†}, Kathleen J. Millen^{*} & Mary E. Hatten

Laboratory of Developmental Neurobiology, The Rockefeller University, 1,230 York Avenue, New York, New York 10021-6399, USA

^{*} These authors contributed equally to this work

In the vertebrate central nervous system (CNS), a cascade of signals that originates in the ectoderm adjacent to the neural tube is propagated by the roof plate to dorsalize the neural tube¹. Here we report that the phenotype of the spontaneous neurological mutant mouse *dreher* (*dr*)^{2–5} results from a failure of the roof plate to develop. Dorsalization of the neural tube is consequently affected: dorsal interneurons in the spinal cord and granule neurons in the cerebellar cortex are lost, and the dorsal vertebral neural arches fail to form. Positional cloning of *dreher* indicates that the LIM homeodomain protein, *Lmx1a*, is affected in three different alleles of *dreher*. *Lmx1a* is expressed in the roof plate along the neuraxis during development of the CNS. Thus, *Lmx1a* is required for development of the roof plate and, in turn, for specification of dorsal cell fates in the CNS and developing vertebrae.

Spontaneously generated mutant mice provide a powerful tool for analysing genetic and epigenetic mechanisms of CNS development⁶. To define further the postnatal phenotype of *dr*¹/*dr*¹ mice, we analysed the cytoarchitecture of the cerebellum and spinal cord. In the cerebellum the vermis was absent in *dr*¹/*dr*¹ mice (Fig. 1a, b) and fewer lobes were present at the midline of *dr*¹/*dr*¹ as compared with wild-type mice (Fig. 1c, d). Although lobe patterning was disrupted, the three layers of the cerebellum were present at the cellular level (Fig. 1d). In ventral areas of the cerebellar system, only the pontine nucleus was reduced in size in *dr*¹/*dr*¹ mice (Fig. 1e, f). In more caudal regions of the CNS, the spinal cord had a flattened appearance and cell density appeared to decrease in the dorsal spinal cord (Fig. 1g, h). Examination of the skeleton of the caudal region of the neural tube revealed that the neural arches failed to fuse on the dorsal midline and bone density appeared to be reduced (Fig. 1i, j).

An intercross between *dr*¹ and wild-type MOLF/Ei mice established a detailed genetic and physical map surrounding the *dreher* locus. From 738 informative meioses, four recombination events defined the *dr*¹ locus to a 0.68-cM interval. The closest proximal and distal markers were *D1Mit109* (0.13 cM) and *D1Mit370* (0.54 cM), respectively. The genetic map of this region is shown in Fig. 2a, b. Although *D1Mit110* and the gene *RXRg* co-segregated with the *dr*¹

[†] Present address: Center for Advanced Biotechnology and Medicine, Department of Neuroscience and Cell Biology, UMDNJ-Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, New Jersey 08854-5638, USA.

locus (Fig. 2a), other mapping studies show that *RXR γ* is not the *dr*³ gene³.

To identify *dr*¹ embryos, we established a second cross between +/*dr*¹ mice that were BL6/C3He for *D1Mit15* and *D1Mit424*. *Dreher* arose on a hybrid inbred strain of C57BL/6J and C3He/FeLe-a/a mice. Importantly, both markers had the C3He/C3He genotype for the 336 progeny that were non-recombinant for distal chromosome 1 (data not shown), indicating that *dr*¹ arose on the C3He chromosome. *D1Mit15* and *D1Mit424* were located on either side of the *dreher* locus and were polymorphic between these strains (Fig. 2a, c). *dr*¹/*dr*¹ embryos were thus genotypically C3He/C3He for both markers, +/+ embryos were BL6/BL6 and heterozygotes were BL6/C3He for both markers (Fig. 2c).

As a first step towards cloning the gene responsible for the *dr* phenotype, we generated yeast artificial chromosome (YAC) and bacterial artificial chromosome (BAC) contigs. We identified YACs

for two simple sequence-length polymorphism (SSLP) markers, *D1Mit110* and *D1Mit370* (for details see <http://www.genome.wi.mit.edu>). Sequence tag site (STS) content mapping showed that *D1Mit109* was on a subset of these YACs and this analysis determined that the YAC contig spanned the *dr*¹ minimal region. Screening a BAC library with *D1Mit110* identified five BACs. By BAC-end sequencing, we found that *RXR γ* (ref. 7) was on the proximal end of the BAC contig (Fig. 3a). Physical mapping of each BAC end showed that 292T7 and 160T7 were the proximal and distal ends of the BAC contig, respectively. Both did not recombine with *dr*¹, showing that the BAC contig was internal to the *dr*¹ minimal region.

To scan the BAC contig for genes, we sequenced 57 random subclones, generating about 70 kilobases (kb) of sequence. BLAST searches revealed no stretches of homology within these sequences. PCR analysis with primers generated to 23 BAC subclones and to each of the BAC ends on genomic DNA from four different *dr* alleles (*dr*^{sst}, *dr*¹, *dr*^{3j} and *dr*^{4j})^{2,3,8} revealed the absence of two STS markers, 160T7 and 53-1, from *dr*^{sst} DNA. Thus, the *dr*^{sst} mutation deletes an internal part of the BAC contig (Fig. 3a).

To search for genes contained in the *dr*^{sst} deletion, we isolated DNA from a wild-type mouse genomic library by screening with clones 160T7 and 53-1. Twenty-five kilobases of DNA was sub-cloned and sequenced. One candidate open reading frame, 160-5, consistently hybridized to human DNA and was used to screen an embryonic day (E)8.5 complementary DNA library. A 1.8-kb cDNA clone was isolated that was identical to expressed sequence tag (EST) AA881470. Genomic analysis showed that only the 3' untranslated region of EST AA881470 was deleted in *dr*^{sst}. Further analysis failed to detect mutations in this gene in the other three alleles (Fig. 3a and data not shown).

BLAST searches of genomic DNA deleted in *dr*^{sst} identified a second candidate gene as a 184-base pair (bp) stretch of near identity with the hamster LIM homeodomain gene, *Lmx1.1* (ref. 9). The coding region of the mouse cDNA was isolated from adult mouse brain by polymerase chain reaction with reverse transcription (RT-PCR). The cDNA shared significant identity throughout the coding region with hamster *Lmx1.1* (94% nucleotide identity, 98% amino-acid identity), showing that this is the mouse homologue, which we have termed *Lmx1a*. Like other LIM homeodomain proteins, it includes two LIM domains and a homeodomain¹⁰. Physical mapping showed that only the 5' half of the gene, which includes the start codon and the two LIM domains, is deleted in *dr*^{sst} (Fig. 3a).

Mutations in the *Lmx1a* gene were then identified in two other alleles of *dreher*. *Dr*^{3j} contains a deletion that includes the ATG and the first LIM domain, preventing the translation of the protein (Fig. 3a). Sequence analysis of *dr*¹/*dr*¹ adult brain cDNA and genomic DNA identified a G-to-A base pair change that alters a conserved cysteine in the LIM1 domain to a tyrosine (Fig. 3b, c). LIM domains are composed of two zinc fingers and this cysteine is involved in coordinating the zinc that is essential for LIM function¹¹⁻¹³. It has also been reported that the LIM1 domain is required for the transcriptional activity of the protein^{9,14}. No other sequence change was observed between +/+ and *dr*¹/*dr*¹.

Northern analysis of *Lmx1a* revealed a transcript of roughly 4.0 kb at E11.5. *In situ* analysis of the cellular distribution of expression showed that the gene is expressed in the roof plate of the CNS (Fig. 3d). In the dorsal aspect of the embryonic spinal cord, *Lmx1a* transcripts were also detected in the neural crest cells on either side of the roof plate, and in the ectoderm in some sections. In the mesencephalon/metencephalon (mes/met) region *Lmx1a* was expressed in the zone adjacent to the roof plate, the rhombic lip, which gives rise to precursors of the cerebellar granule neurons. Expression was also seen in the posterior aspect of the developing cortex, where there was a broad band of *Lmx1a* transcripts. Thus, *Lmx1a* is localized to the roof plate and immediately surrounding

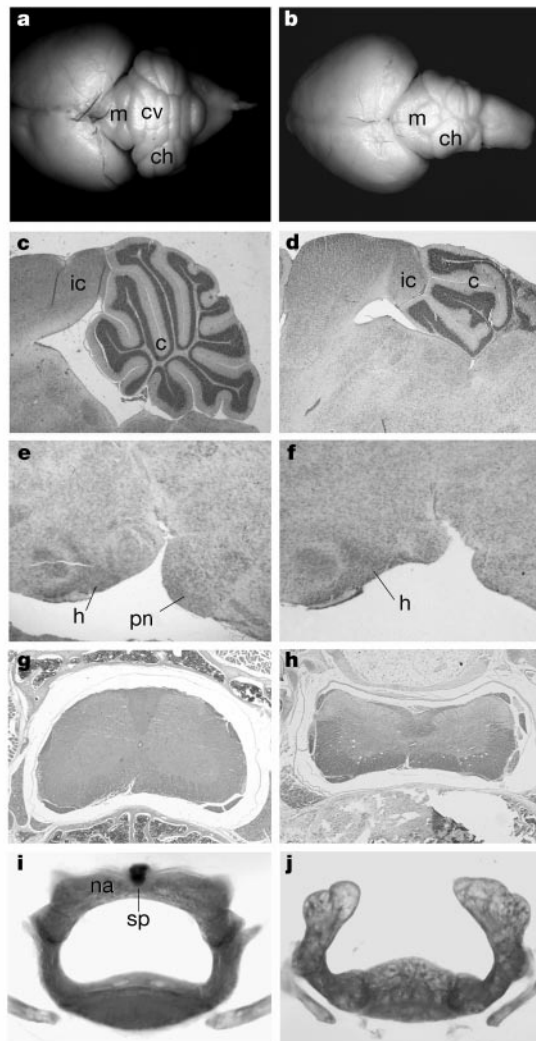
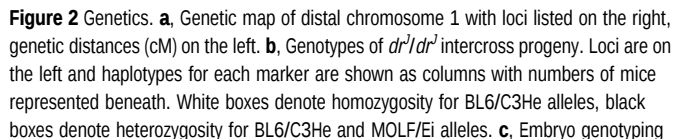


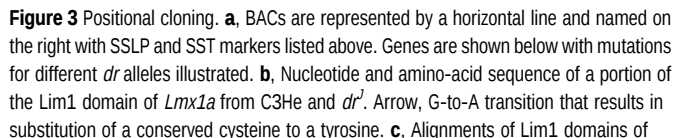
Figure 1 Postnatal brain phenotype. Postnatal day (P)18–20 +/+ (a, c, e, g, i) and *dr*¹/*dr*¹ (b, d, f, h, j) specimens. In dorsal views (a, b), cerebellar vermis (cv) and hemispheres (ch) are denoted. c–f, Nissl stained mid-sagittal sections of cerebellum and dorsal midbrain (c, d) and ventral mid-hindbrain (e, f). Fewer numbers of vermal lobes were seen in *dr*¹ and the size of the pontine nucleus (pn) was reduced. g, h, Nissl stained transverse sections of L2 spinal cord; i, j, transverse views of T13 vertebrae stained with alcian red/alizarin blue. Note the abnormal morphology of the spinal cord and the absence of the vertebral neural arches (na) and spinous process (sp) on the dorsal midline. c, cerebellar anlage; ic, inferior colliculus; h, hypothalamus; m, midbrain.

To investigate the role of dr^f in embryonic development, we focused on the caudal aspect of the CNS and the mid-hindbrain region. Transverse sections through the caudal portion of the neural tube of E11.5 embryos revealed a thickening of the dorsal midline (see below), indicating a possible defect in the roof plate. To visualize roof-plate cells, we localized the roof-plate marker MafB, a basic domain–leucine-zipper transcription factor encoded by the *Kreisler* locus^{15,16}, which is expressed in the roof plate¹⁷. By immunocytochemistry, MafB antibody staining of E9.75 embryos detected roof-plate cells localized at the dorsal aspect of the

To examine whether the roof plate formed at later stages of



strategy. C57BL6/J and C3He/FeLe-a/a chromosomes for $+d^r$ animals that have the BL6/C3He genotype for both *D1Mit15* and *D1Mit424* are illustrated with genotypes for both SSLP markers of offspring shown in tabular form and PCR polymorphisms adjacent (relative molecular mass ($\times 10^{-3}$) on the left. B, C57BL6/J; C, C3He/FeLe-a/a; B/C; C57BL6/J and C3He/FeLe-a/a.



766

development, we immunostained transverse sections of E11.5 with antibodies against MafB. As seen at E9.75, no roof-plate cells were detected (data not shown) in dr^j/dr^j as compared with $+/+$ embryos. We then examined expression of the gene for bone morphogenetic protein 6 (*Bmp6*)¹⁷, which is restricted to the roof plate in the dorsal

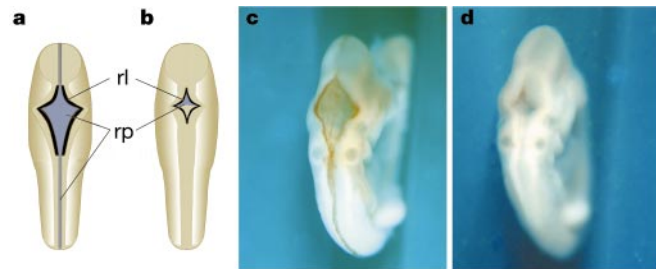


Figure 4 dr^j/dr^j embryos lack a roof plate. Schematic representation of $+/+$ (a) and dr^j/dr^j (b) E9.75 embryos. At the mes/met border (top), roof plate (rp) is shown in blue-grey. The rhombic lip (rl) is shown in black. c, d, Whole-mount immunostaining of E9.75 embryos with anti-MafB antibody staining revealing the roof plate along the dorsal aspect of the spinal cord and midbrain and across the roof plate at the mes/met border in $+/+$ (c). In dr^j/dr^j embryos (d), no roof-plate cells are detected in the spinal cord or midbrain. A severely reduced roof plate is seen at the level of the presumptive cerebellar territory.

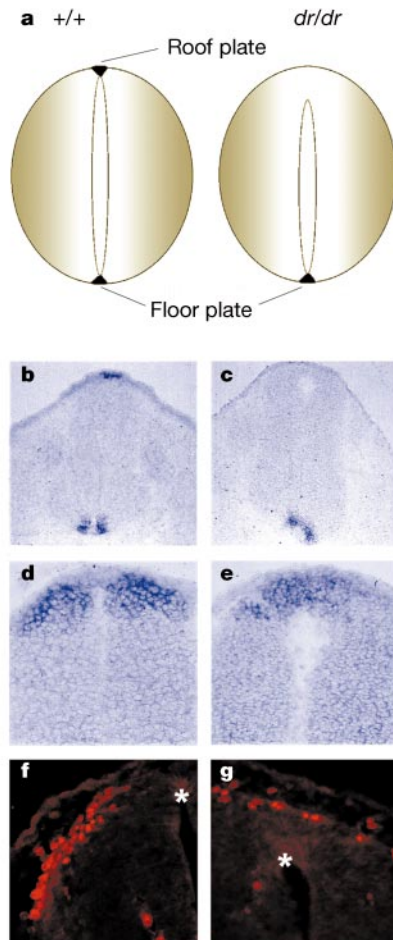


Figure 5 Spinal cord phenotype. a, Schematic representation of wild-type and dr^j/dr^j E11.5 spinal cord showing roof plate and floor plate. b, *Bmp6* is expressed in $+/+$ roof plate but not in dorsal dr^j/dr^j spinal cord (c), although ventral *Bmp6* expression is maintained. *Math1* is detected in two zones adjacent to the roof plate in $+/+$ (d) embryos. By contrast, in dr^j/dr^j (e) there is a contiguous dorsal zone of expression. Localization of the D1 dorsal interneuron population, detected by immunocytochemistry to the Lim homeodomain proteins LH2A and LH2B, reveals fewer interneurons in dr^j/dr^j (g) compared with $+/+$ (f). Asterisk, midline.

aspect of the neural tube and to cells adjacent to the floor plate in the ventral region. In dr^j embryos, expression was not detected in the dorsal aspect of the neural tube, suggesting an absence of roof-plate cells. Normal expression was seen in the ventral region (Fig. 5b, c). We then investigated the pattern of expression of *Math1* (ref. 18), which is normally expressed in the ventricular zone adjacent to the roof plate. In dr^j/dr^j mutants, we observed *Math1* transcripts in a contiguous zone of cells across the dorsal region of the neural tube (Fig. 5d, e), indicating that the two dorsal domains had fused in the absence of the roof plate. These data show that dr^j/dr^j embryos lack a roof plate in the caudal portion of the neural tube.

The absence of a roof plate in the caudal neural tube prompted us to examine dorsal cell specification in the spinal cord. Previous studies have shown that the specification of dorsal spinal cord interneurons is dependent on the roof plate¹. In the dorsal aspect of the developing spinal cord, one population of dorsal interneurons (D1) is defined by expression of LIM homeodomain proteins that are recognized by an antibody against LH2 (refs 17, 19). These cells originate dorsally, just lateral to the midline in $+/+$ embryos. In dr^j/dr^j embryos at E11.5, the number of LH2-positive cells appeared to be reduced, and stained cells were situated on the dorsal midline (Fig. 5f, g). Thus, dorsal neural specification was disrupted by the dr^j locus in the developing spinal cord.

In view of the role of dr^j in the caudal neural tube, and of recent experiments²⁰ extending the general model of dorsal cell specifica-

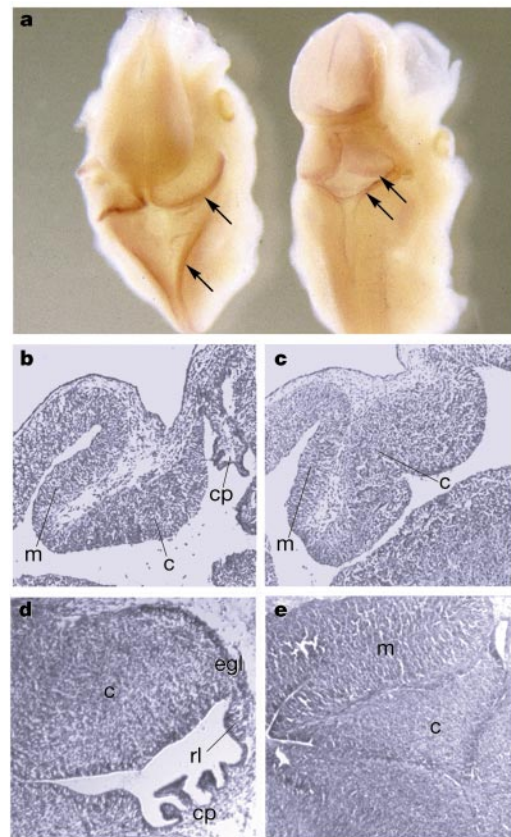


Figure 6 Cerebellar development. a, *Math1* whole-mount *in situ* hybridization in $+/+$ (left) and dr^j/dr^j (right) at E11.5 show a severe reduction in the rhombic lip (arrows show the anterior and posterior sides) bordering the roof plate across the fourth ventricle. Nissl-stained sagittal sections of E12.5 $+/+$ (b) and dr^j/dr^j (c) embryos show an absence of choroid plexus (cp). d, e, Nissl-stained sagittal sections of E14.5 $+/+$ (d) embryos show the rhombic lip (rl) with granule-cell progenitors (egl) migrating from the rl across the roof of the anlage (c). In dr^j/dr^j embryos (e), the choroid plexus, rhombic lip and granule-cell progenitors were missing in many sections. The cerebellar anlage was also reduced in size. m, midbrain.

tion to the cerebellar granule cell, we examined the phenotype of dr^j/dr^j embryos at the mes/met border where the cerebellar territory arises. In this region, the rhombic lip is adjacent to the overlying roof plate. Whole-mount, *in situ* hybridization with *Math1* revealed that the roof plate and rhombic lip were significantly reduced in dr^j/dr^j embryos at E11.5 (Fig. 6a, arrows). In the developing cerebellar anlage at E12.5, the choroid plexus, a derivative of the roof plate (R. Wingate, personal communication), was missing in a number of sagittal sections in mutant embryos (Fig. 6b, c). In addition, at E14.5 the rhombic lip, which is the source of granule cell progenitors^{20,21}, was absent in many sections of dr^j/dr^j embryos (Fig. 6d, e). Thus, the roof plate of the met/mes region and the surrounding rhombic lip of the cerebellar territory were significantly reduced in the dr^j/dr^j embryo. These data indicate that, as in more caudal regions of the embryo, the roof plate may be essential for dorsal patterning within the cerebellum.

Our data underscore the role of the roof plate in normal CNS development, and offer new tools with which to analyse the mechanism of specification of roof-plate cells. Whereas previous studies on the role of the roof plate have focused on the spinal cord, our studies indicate that it might be possible to extend the general model of dorsal fate specification, mediated by factors transmitted by the roof-plate cells, to rostral areas of the developing nervous system. A clearer understanding of roof-plate cell specification, and of the role of these cells in propagating locally derived signals, will be critical to deciphering the mechanisms underlying the specification of dorsal cell fates in the CNS. □

Methods

Mice, crosses and DNA

Dreher⁺ heterozygous mice (B6C3Fe-a/a- dr^j), C57BL/6J, C3HeB/FeLe-a/a and *Mus musculus molossinus* MOLF/Ei mice were purchased from the Jackson Laboratory. The genetic map was based on an intersubspecific intercross between dr^j and MOLF/Ei. Of 1,702 postnatal F2 mice generated, 369 were identified as dr^j/dr^j by their morphological and neurological phenotypes. Tail DNA was prepared from all 369 dr^j/dr^j F2 progeny and genotyped by PCR²² with *D1Mit424* and *D1Mit15*. Recombinants were further genotyped with the markers listed in Fig. 2. For embryonic analysis, adult $+/dr^j$ mice heterozygous for the C57BL/6J and C3HeB/FeLe-a/a alleles of both *D1Mit424* and *D1Mit15* were intercrossed. The genotypes of *D1Mit424* and *D1Mit15* identified offspring as $+/+$, $+/dr^j$ or dr^j/dr^j as described (Fig. 2c) from yolk sac DNA. Genomic DNA from the dr^j parental strains was prepared as described²². Genomic DNA from the dr^{3j} , dr^{4j} , parental strains (C57BL/6J, C3H/HeSnJ and CBA/J) and normal littermate controls was purchased from the Jackson Laboratory. Genomic DNA from the dr^{sst} strain of mice was obtained from fixed tissue preserved in 70% ethanol, provided by D. Wahlsten. After rehydration, DNA was extracted. Parental strain DNA for dr^{sst} (C57BL/6J and BALB/cWahl) was prepared from fixed tissue and frozen liver samples.

Phenotypic analysis

Adult brain dr^j tissue was fixed in Bouin's fixative and processed for paraffin sections as described²². For histology, embryos were fresh frozen, sectioned and Nissl stained. For immunohistochemistry, tissue was processed as described²¹ using LH2 and MaB antibodies (provided by T. M. Jessell). We conducted digoxigenin *in situ* hybridization²³ with a *Bmp6* probe provided by B. Hogan and a *Math1* probe provided by J. Johnson. Radioactive *in situ* hybridization²⁴ was conducted with a *Lmx1a* probe spanning the first 263 bp of coding sequence. Whole-mount *in situ* hybridizations and postnatal skeletal staining were conducted as described²⁵. Analysis of dr^{3j} , dr^{4j} and dr^{sst} was not possible, as these strains are extinct.

Identification of the *Lmx1a* gene

D1Mit110-positive BACs were purchased from Research Genetics. Sequence from BAC ends allowed generation of STS markers to map each end relative to the other BACs. 292G21 and 160F9 were cloned into pBluescript (Stratagene) and random *HindIII* fragments from clones were sequenced using TaqFS chemistry with Big Dye terminators (ABI) on ABI 377XL sequencing apparatus. Sequence was analysed using Lasergene (DNASTAR) software. STS markers were designed and mapped on the BAC contig and genomic DNA. Multiple sets of primers derived from 160T7, an end clone from BAC 160F9, and 53-1, a 292G21 subclone, failed to amplify DNA from dr^{sst} , although the expected products were amplified from all other strains.

We screened a 129SvJ genomic library (Stratagene) with BAC subclones 160T7 and 53-1. Positive clones were restriction mapped, subcloned and sequenced. Subclones with large open reading frames were hybridized at low stringency to human genomic DNA (Clontech). Northern analysis was conducted with RNA isolated from embryonic and

adult wild-type and dr^j tissue. An E8.5 cDNA library (B. Hogan) was screened with 160-5, a 2.1-kb phage subclone containing an exon of EST A4881470. Partial cDNAs for this transcript were isolated from dr^j and parental strains and sequenced. No differences were observed. Southern analysis of BAC DNA mapped the 3' end of this gene to the BAC contig. Multiple sets of primers from intron sequences at the 3' end of this gene (for example, forward: GTACGGACAATAAAGGACGAG, reverse: GACTGATGCTGAAGC-CATGAG) failed to amplify genomic DNA from dr^{sst} DNA, although the predicted products were amplified from all other strains.

The coding region of the mouse *Lmx1a* gene was cloned from adult brain cDNA from C57BL/6J, C3HeB/FeLe-a/a and dr^j strains using overlapping sets of primers based on the hamster *Lmx1.1* gene (GenBank X81406). To amplify specifically the Lim1 domain-containing exon of *Lmx1a*, we designed flanking primers based on flanking intron sequences (forward: GGCAACATCTGTTGCTGTTG, reverse: GAAGCAGGCCTACTTCTC). These primers failed to amplify dr^{3j} and dr^{sst} DNA. Primers within the ATG-containing exon (forward: ATGTTGGACGGCCTGAAG, reverse: GAGGTCTCAATCGC ACTTTG) also failed to amplify dr^{3j} and dr^{sst} DNA. In contrast, primers derived from the intron 3' to the Lim1 domain-containing exon (forward: GTCTTTCTAAGCATCTGT-TAC, reverse: GTGACAGTAACCAGAATCAC), amplified a 77-bp fragment from dr^{3j} and all other strains, but not dr^{sst} . Primers from the 3' end of the coding region of *Lmx1a* (forward: GCAGTCCAGAGTGGGAAAC, reverse: CTCAGAGGTGAATAGGAATTC) amplified a 68-bp fragment from both dr^{3j} and dr^{sst} , and all parental strains.

The deletion of the ATG-containing exon and the Lim1 domain-containing exon was confirmed by Southern analysis of dr^{3j} DNA using a probe spanning the first 263 bp of the coding region. No differences were detected in dr^j DNA by extensive genomic Southern analysis using most probes derived in this study. The point mutation in dr^j was confirmed by sequencing clones derived from cDNA and genomic DNA. No polymorphisms in the entire coding sequence were observed in C3HeB/FeLe-a/a or C57BL/6 cDNA or genomic DNA. The *Lmx1a* transcript was detected on Multiple Tissue poly(A)⁺ northern blots (Clontech).

Received 9 September 1999; accepted 4 January 2000.

- Lee, K. & Jessell, T. The specification of dorsal cell fates in the vertebrate central nervous system. *Annu. Rev. Neurosci.* **22**, 261–294 (1999).
- Lyons, M., Rastan, S. & Brown, S. *Genetic Variants of the Laboratory Mouse* 195 (Oxford Univ. Press, Oxford, 1996).
- Bergstrom, D., Gagnon, L. & Eicher, E. Genetic and physical mapping of the dreher locus on mouse chromosome 1. *Genomics* **59**, 291–299 (1999).
- Seikiguchi, M., Shimai, K., Guo, H. & Nowakowski, R. Cytoarchitectonic abnormalities in hippocampal formation and cerebellum of dreher mutant. *Brain Res. Dev. Brain Res.* **67**, 105–112 (1992).
- Seikiguchi, M. *et al.* Disruption of neuronal migration in the neocortex of the dreher mutant mouse. *Brain Res. Dev. Brain Res.* **77**, 37–43 (1994).
- Sidman, R. in *Genetics of Neurological and Psychiatric Disorders* (eds Kety, S., Rowland, L., Sidman, R. & Matthysse, S.) 19–46 (Raven, New York, 1983).
- Almasan, A., Mangelsdorf, D., Ong, E., Wahl, G. & Evans, R. Chromosomal location of the human retinoid X receptors. *Genomics* **20**, 397–403 (1994).
- Walsten, D., Lyons, J. & Zagaja, S. Shaker short-tail, a spontaneous neurological mutant in the mouse. *J. Hered.* **74**, 421–425 (1983).
- German, M., Wang, J., Chadwick, R. & Rutter, W. Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin mini-enhancer complex. *Genes Dev.* **6**, 2165–2176 (1992).
- Schmeichel, K. & Beckerle, M. The LIM domain is a modular protein-binding interface. *Cell* **79**, 211–219 (1994).
- Perez-Alvarado, G. *et al.* Structure of the carboxy-terminal LIM domain from the cysteine rich protein CRP. *Nature Struct. Biol.* **1**, 388–398 (1994).
- Perez-Alvarado, G. *et al.* Structure of the cysteine-rich intestinal protein, CRIP. *J. Mol. Biol.* **257**, 153–174 (1996).
- Konrat, R., Weiskitchen, R., Krautler, B. & Bister, K. Solution structure of the carboxy-terminal LIM domain from quail cysteine-rich protein CRP-2. *J. Biol. Chem.* **272**, 12001–12007 (1997).
- Johnson, J. *et al.* Transcriptional synergy between LIM-Homeodomain proteins and basic helix-loop-helix proteins: the LIM2 domain determines specificity. *Mol. Cell Biol.* **17**, 3488–3496 (1997).
- Cordes, S. & Barsh, G. The mouse segmentation gene *kr* encodes a novel basic domain-leucine zipper transcription factor. *Cell* **79**, 1025–1034 (1994).
- Kataoka, K., Noda, M. & Nishizawa, M. Maf nuclear protein recognizes sequences related to an AP-1 site and forms heterodimers with both Fos and Jun. *Mol. Cell Biol.* **14**, 700–712 (1994).
- Liem, K., Tremml, G. & Hatten, M. The role of roof plate and its resident TGF β -related proteins in neuronal patterning in the dorsal spinal cord. *Cell* **91**, 127–138 (1997).
- Ben-Arie, N. *et al.* *Math1* is essential for genesis of cerebellar granule neurons. *Nature* **390**, 169–172 (1997).
- Liem, K. F. Jr, Tremml, G., Roelink, H. & Jessell, T. M. Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. *Cell* **82**, 969–979 (1995).
- Alder, J., Cho, N. & Hatten, M. Embryonic precursor cells from the rhombic lip are specified to a cerebellar granule neuron identity. *Neuron* **17**, 389–399 (1996).
- Alder, J., Lee, K. J., Jessell, T. M. & Hatten, M. E. Generation of cerebellar granule neurons *in vivo* by transplantation of BMP-treated neural progenitor cells. *Nature Neurosci.* **2**, 535–540 (1999).
- Millonig, J., Millen, K. & Hatten, M. A detailed genetic map of the weaver locus on Chromosome 16. *Mamm. Genome* **7**, 616–618 (1996).
- Schaeren-Wiemers, N. & Gerfin-Moser, A. A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: *in situ* hybridization using digoxigenin-labelled cRNA probes. *Histochemistry* **100**, 431–440 (1993).
- Millen, K. & Hui, C.-C. in *A Laboratory Guide to RNA: Isolation, Analysis and Synthesis* (eds Millen, K. & Hui, C.-C.) 339–355 (Wiley-Liss, New York, 1996).
- Parr, B., Shey, M., Vassieva, G. & McMahon, A. Mouse Wnt genes exhibit discrete domains of expression in the early embryonic CNS and limb buds. *Development* **119**, 247–261 (1993).

Acknowledgements

We thank D. Wahlsten for *dr^{wt}* embryos; J. Johnson for the *Math1 in situ* probe; T. Jessell for antibodies against LH2 and MafB; B. Hogan for the *Bmp6* probe and E8.5 cDNA library. M. Cooper and D. Patterson provided technical assistance; C. Bonal and L. Chemes participated in the sequence analysis of *dr^l*; N. Adams provided assistance with digital imaging; and D. Benyaklef prepared the Figures. We thank N. Heintz, T. Jessell, J. Alder, K. Lee, P. Matteson, R. Slavkov, R. Wingate and K. Zimmerman for helpful discussions. This work was supported by a Program Project grant (M.E.H.)

Correspondence and requests for materials should be addressed to M.E.H. (e-mail: hatten@rockvax.rockefeller.edu). GenBank accession numbers for murine *Lmx1a* and *EST AA881470* are AF226662 and AF226663 respectively.

Forebrain peptides modulate sexually polymorphic vocal circuitry

James L. Goodson* & Andrew H. Bass*†

* Department of Neurobiology and Behaviour, Cornell University, Ithaca, New York 14853, USA

† Bodega Marine Laboratory, University of California, Bodega Bay, California 94923, USA

The peptide arginine-vasopressin (mammals) and its evolutionary precursor arginine-vasotocin (non-mammals) modulate reproductive physiology and numerous related social behaviours, as do oxytocin (mammals) and its homologues mesotocin and isotocin (fish)¹. The distributions in the brain and/or the behavioural functions of these peptides often differ between the sexes^{2–4}, and between species with divergent social structures^{3,5,6}. Here we present neurophysiological evidence that males with vocal characteristics typical of females share a pattern of neuro-peptide function with females rather than conspecific males. The plainfin midshipman fish (*Porichthys notatus*) has two male morphs with different reproductive tactics and vocalizations (a key species-typical behaviour which varies in its physical attributes and contextual usage, depending on the morph's social strategy)^{7,8}. Forebrain-evoked, rhythmic vocal-motor activity that precisely mimics natural vocalizations was modulated by arginine-vasotocin, isotocin and their antagonists delivered to the preoptic area–anterior hypothalamus, a primary site for behavioural integration in all vertebrates. Peptides had different

effects in males that acoustically court females (arginine-vasotocin-sensitive) than in females and sneak-spawning males (isotocin-sensitive), showing that the neuromodulatory mechanisms that establish reproduction-related behaviour can be dissociated from gonadal sex.

Each of the three reproductive midshipman phenotypes follows a fixed developmental path and is characterized by a distinct suite of behavioural, somatic, neural and endocrine traits⁸. Type I males excavate and defend a nest under a rock, from which they generate a long advertisement 'hum' (> 1 min). Investigating females deposit eggs on the roof of the nest and immediately depart; the type I male provides all care for the eggs and hatchlings. Type II males are similar to females in body size and shape and, unlike type I males, do not engage in nest construction, courtship or parental care. These males compete for fertilizations by ejecting sperm from the periphery of a type I nest (satellite spawning) or by entering a nest (sneak spawning). All adult morphs emit brief (< 250 ms) agonistic 'grunts', although only type I males emit grunts in long trains. Type II sneaker males and females generate isolated, low-amplitude grunts. Type I males use hums for courtship and grunts in the defence of the nest site, eggs and hatchlings, whereas agonistic grunts in type II males and females have been identified only in non-reproductive contexts⁷. Hence, the differential use of vocalization is strongly tied to the differential expression of parental behaviour, nest defence and courtship—all behaviours that are modulated by peptides of the vasopressin–oxytocin family in other vertebrate species.

We specifically tested the hypothesis that arginine-vasotocin (AVT) and isotocin (IT) delivered into the preoptic area–anterior hypothalamus (POA–AH) would differentially modulate vocalization in the different adult midshipman morphs. This hypothesis was based on two observations. The POA–AH of all vertebrates contains a high density of vasopressin–oxytocin elements and is a primary site of complex behavioural and physiological integration (including vocalization)^{4,9,10}; and variability in preoptic AVT elements is associated with sex change and divergence of reproductive tactics in teleost fish^{2,11,12}.

Midshipman vocalization can be readily investigated in a neuro-physiological recording paradigm, as output recorded from the hindbrain vocal-motor circuitry and acoustic production are related in a one-to-one fashion (Fig. 1a)¹³. Intracellular recording and staining show that the temporal properties of a pacemaker-sonic motor neuron (PN–SMN) circuit's rhythmic volley directly determine the physical attributes of vocalizations. Hence, if the PN–SMN circuit is firing at 100 Hz, the fundamental frequencies of sonic muscle contraction and of the resultant vocalization are both

Table 1 Peptide effects on vocal-motor response to POA–AH stimulation 2 and 30 min post-treatment

	2 min			30 min		
	Summed burst duration	Mean burst duration	No. of bursts	Summed burst duration	Mean burst duration	No. of bursts
Type I males						
anti-AVP	n.s.	n.s.	n.s.	$r = 0.685$, $P = 0.052$	n.s.	n.s.
AVT	$r = -0.650$, $P = 0.024$	$r = -0.670$, $P = 0.020$	n.s.	n.s.	n.s.	n.s.
IT	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
anti-OXY	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Females						
anti-OXY	$*P = 0.049$	n.s.	$*P = 0.036$	n.s.	n.s.	n.s.
IT	n.s.	n.s.	n.s.	$r = -0.729$, $P = 0.011$	$r = -0.658$, $P = 0.026$	$r = -0.617$, $P = 0.032$
AVT	n.s.	n.s.	n.s.	n.s.	n.s.	$*P = 0.046$
anti-AVP	n.s.	n.s.	$*P = 0.036$	n.s.	n.s.	n.s.
Type II males						
anti-OXY	$*P = 0.049$	n.s.	$*P = 0.046$	n.s.	$*P = 0.049$	n.s.
IT	n.s.	n.s.	n.s.	$r = -0.550$, $P = 0.056$	n.s.	$r = -0.578$, $P = 0.045$
AVT	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
anti-AVP	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

Spearman rank correlations for multiple-group data are shown with *r*-values. Mann-Whitney *P*-values for two-group comparisons are denoted with an asterisk. See Fig. 2 for details on doses, subject *ns* and statistics.

100 Hz. This allows us to recognize fictive social vocalization on a physiological level. Descending brainstem connections to the medullary pattern-generating circuit have been identified¹⁴ and naturalistic vocal-motor activity can be elicited by electrical stimulation of numerous neural loci, including the POA–AH (Fig. 1). The midshipman vocal-motor system exhibits many evolutionarily conserved characteristics, including a PN–SMN circuit that is considered to share embryonic origins with vocalization circuitry in tetrapods¹⁵.

We stimulated the POA–AH through a tungsten electrode and simultaneously recorded vocal-motor output from the hindbrain through an extracellular electrode placed on an occipital nerve root (comparable to the hypoglossal nerve of tetrapods¹⁵). A representative fictive grunt produced by POA–AH stimulation is shown in Fig. 1a and a marked stimulation site located in the ventral tuberal region of the anterior hypothalamus is shown in Fig. 1b (the site is consistent with those marked in 22 other subjects; all morphs represented). Preoptic neurons that are immunoreactive for AVT and IT in teleosts produce a dense fibre innervation of the tuberal hypothalamus and other regions of the POA–AH¹⁶.

We measured the baseline response by providing stimulus trains at 1-s intervals (see Methods). This was followed by pressure delivery of a neuropeptide, an antagonist or vehicle control (the broad dosage range was selected to include concentrations that are effective in modulating behaviour^{17–19} and reproduction-related neurophysiology²⁰ in other vertebrate species). The vocal-motor response to POA–AH stimulation was then reassessed at multiple time points after treatment (including 2, 10 and 30 min for all subjects). The time required for recovery from peptide treatments varied and ranged from 30 min to 2 h. The effects of arginine-vasopressin (AVP) antagonists were often slightly delayed (probably because of time required for endogenous peptide displacement), whereas responses to other treatments were usually rapid. The data shown in Fig. 2 thus focus on the 10-min time point, as all effective substances produced strong effects at this time. Statistical results from the 2-min and 30-min time intervals are given in Table 1. For maximal clarity, statistical comparisons for multiple dosage levels ($\bullet$ 2 groups) are limited to Spearman rank correlations (corrected for ties; subject *n*s per dosage are the number of data points entered per dosage). As shown in Fig. 2 and Table 1, this revealed clear and unidirectional dose-dependent effects; pairwise comparisons were therefore deemed excessive given the number of treatments and dependent measures. We used Mann–Whitney *U* tests for unpaired two-group comparisons.

Fictive vocalization in type I males was inhibited by AVT and facilitated by an AVP V₁ receptor antagonist (($\bullet$ -mercapto- $\bullet$, $\bullet$ -cyclopenta-methylenepropionyl¹, O-Me-Tyr², Arg³)-Vasopressin; Fig. 2 and Table 1). AVT produced significant dose-dependent decreases in total vocal-motor response (summed durations of all bursts elicited). This effect on total response was produced by modulations of both burst duration (Table 1) and the number of vocal-motor bursts elicited (Fig. 2). The antagonist produced significant dose-dependent facilitations of total response and burst duration measures, but not of the number of bursts, indicating that AVT may act primarily to modulate the duration of vocal bursts in type I males. Most vocal-motor responses elicited were grunt-like, and longer duration output ($\bullet$ 250 ms) was not obtained in enough subjects for separate analysis. However, given that courtship hums and grunts are primarily distinguished from each other on the basis of duration⁷, AVT modulation of burst duration may serve to set vocal-motor parameters (vocalization type) appropriate to a given social context.

The effects of AVT were highly specific, as no effects were observed after delivery of the structurally similar IT or a selective oxytocin (OXY) antagonist ((des-glycinamide⁹, d(CH₂)₅¹, O-Me-Tyr², Thr⁴, Orn⁸)-Vasotocin) which has a very weak affinity for the mammalian V₁ receptor²¹ (Fig. 2 and Table 1). These results are consistent with findings that the teleost AVT receptor is almost completely insensitive to IT²². In addition, the effects of AVT were probably due to specific action in the POA–AH and not diffusion to other sites, as delivery of 1,000 ng AVT to the ventral telencephalon of two type I males produced no observable effects on vocalization elicited by POA–AH stimulation (data not shown). The teleost ventral telencephalon is innervated by AVT-immunoreactive fibres¹⁶ and contains nuclei considered to be homologous to the septal and amygdalar nuclei of tetrapods²³. This is probably the only nearby region where AVT could have produced effects on vocalization.

In sharp contrast to type I males, females responded to IT with significant dose-dependent decreases on all measures (Fig. 2 and Table 1). Effects on total response were most strongly associated with modulations of the number of bursts elicited, and these parameters were facilitated by the OXY antagonist. In contrast, a high dose of AVT (1,000 ng) produced only a modest but significant facilitation of the number of bursts elicited at 10 and 30 min after treatment (Fig. 2 and Table 1). This dose was sufficient to abolish vocal-motor output in two of three type I males (Fig. 2). The limited effects of AVT in females could be due to binding at either AVT or IT receptors, as the isotocin receptor is AVT-sensitive at concentrations

~3.5 times that of IT²⁴. The relative strengths of the effects of AVT and IT on the female vocal-motor response are consistent with this differential agonist selectivity of the IT receptor. Alternatively, the AVT receptor could have a limited role in female vocalization, as treatment with the AVP antagonist produced a very modest but significant facilitation of the number of bursts elicited at 2 min after treatment (Table 1). However, the AVP antagonist used has a low affinity for the OXY receptor²¹ and could potentially produce weak effects by binding to the IT receptor.

Type II males exhibited a clear pattern of results which was nearly identical to that obtained in females. Administration of IT produced significant dose-dependent inhibitions of both total response and number of bursts, and the OXY antagonist produced significant and occasionally massive facilitations of these measures (Fig. 2 and

Table 1). The OXY antagonist also had a delayed effect on burst duration (Table 1). Unlike type I males, type II males were not sensitive to AVT or the AVP antagonist (Fig. 2 and Table 1; but note AVT range overlap with females).

In summary, we have shown that intrasexual behavioural divergence is associated with variability in AVT and IT function, and that this divergence in AVT and IT function between males is associated with one male phenotype expressing a profile of neuropeptide function typical of conspecific females. Thus, fictive vocalization is inhibited by AVT in type I males, whereas females and type II sneaker males are more sensitive to IT. Overall, peptide effects in type II males and females were most strongly associated with modulations of the number of bursts, whereas the effects in type I males were mainly associated with modulation of burst duration.

Figure 2 Vocal-motor responses to POA–AH stimulation (top, type I males; middle, females; bottom, type II males), 10 min after delivery of AVT, IT, their respective antagonists (anti-AVP; anti-OXY), or vehicle control (CON) into the stimulation site. Each panel contains three superimposed graphs, colour-coded to three response parameters. Median response for each parameter is given as a percentage of the baseline (post/pre). Minimum and maximum scores are given below each column, colour-coded to the

relevant parameter. Significant Mann–Whitney comparisons (CON and a single peptide dosage) are denoted by asterisks. Significant Spearman rank correlations (for multiple dosages) are shown with colour-coded lines extended above the relevant doses, including CON as a 0 ng dose. Order of dose amounts (left to right) is reversed for peptides and their antagonists.

This observation is consistent with the natural behaviour of these fish, as only type I males exhibit multiple vocalization types of different durations⁷.

Our experiments provide evidence that central IT and AVT modulate behaviour in fishes, and that, as in mammals^{3,25}, sex differences in the potency of both AVP and OXY homologues exist in a non-mammalian vertebrate. Our findings are also consistent with findings that peptides of the vasopressin–oxytocin family modulate behaviour differently between avian and rodent species that differ in social structure^{3,5,6,17}, and with correlational anatomical findings that suggest a role for AVT/AVP in the production of intraspecific behavioural divergence, including vocalization^{2,12,26–29}. Our experiments now provide direct neurophysiological evidence that intrasexually differentiated behaviour is modulated divergently by forebrain neuropeptides. AVT/AVP is known to modulate vocalization in birds¹⁷, anuran amphibians⁴ and mammals³⁰; our results explicitly show that this may occur by specific action in the POA–AH, a region that is implicated in vocal production across many vertebrate taxa^{4,19}. Similar data exist for OXY function in mammals¹⁹. The modulation of vocal-burst duration in the POA–AH of type I males by AVT is also the only evidence to date, to our knowledge, that motor patterning may be dynamically influenced by neurochemicals in this forebrain region. Our experiments show that males that express female-typical patterns of behaviour may also share a pattern of forebrain neuromodulation in common with females rather than conspecific males. Given that the POA–AH regulates both social behaviour and the pituitary–gonadal axis, these findings provide strong evidence that gonadal sex and social/reproductive tactics may be uncoupled from each other and regulated independently within the same brain region. Thus, reproduction-related characteristics may be selected upon as dissociable units, allowing an evolutionarily labile patterning of sexual behaviours which is not limited by gonadal sex. M

Methods

Collection, housing, and recording procedures are described fully elsewhere^{13,14} and are in accordance with institutional guidelines of Cornell University.

Neurophysiological recordings

Dorsal craniotomy was conducted under general anaesthesia (immersion in 0.2% benzocaine). The subject was placed in a plexiglass tank and the skull stabilized with a chronic headholder. The brain was kept covered with a fluorocarbon (Fluorinert, 3M). Fish were perfused through the mouth with salt water at 15–16°C. During recording, pancuronium bromide (0.5 mg kg⁻¹; Astra) was used for immobilization and fentanyl (1 mg kg⁻¹; Sigma) was used for analgesia.

We used tungsten electrodes, insulated except at their tips (125 µm diameter; 20 mm exposed tips; 5 MΩ impedance), to deliver brief (30–35 ms) trains of low-amplitude stimuli (0.05–0.1 ms duration, 280–350 Hz). We monitored the output from an occipital root unilaterally with a teflon-coated, bare silver-ball electrode (50–100 µm diameter). Both sides of the brain fire in phase, so unilateral monitoring reflects bilateral synchrony of the vocal circuit¹³. Recordings were digitized using a Power Macintosh 8100 with IGOR Pro software (Wavemetrics; software customized by R. Wytenbach, Cornell University).

Consistent vocal-motor response to POA–AH stimulation was obtained in 51 subjects (mean treatments per fish, 1.4; maximum, 3; 18 type I males, 13.7–21.0 cm; 17 females, 12.5–15.1 cm; 16 type II males, 8.7–12.3 cm). Subjects were collected between April and August of 1998 and 1999; experiments were conducted in late spring through early fall (September through early November, 1998; late April through early November, 1999). Once consistent output was established, baseline response was assessed by delivering 15 trains of stimuli at 1-s intervals; baseline response was defined as vocal-motor activity elicited during the 15 s of stimulation and the subsequent 45 s.

Peptide delivery

After baseline establishment, the tungsten electrode was removed and a glass pipette (tip diameter ~20 µm) was inserted for pressure delivery of peptide, an antagonist or vehicle control (teleost ringer). Substances were delivered in 0.2 ml vehicle over a 10–15-s period. The tungsten electrode was immediately re-inserted for collection of post-treatment data. Accurate placement in the previous site was readily achieved using micromanipulator coordinates and multiple surface landmarks. For subjects used in multiple experiments, a minimum of 2 h was allowed between treatments. Peptides were obtained from Bachem California and Sigma.

Site marking

Stimulation sites were marked in 23 fish by inducing small electrolytic lesions ($n = 11$), pressure injection of india ink ($n = 8$) or iontophoretic application of biotin compound tract tracers (10% dextran biotin or 5% neurobiotin in 3 M KCl; $n = 4$). Standard histological and reaction procedures were used¹⁴.

Received 7 September; accepted 25 November 1999.

- Moore, F. L. Evolutionary precedents for behavioral actions of oxytocin and vasopressin. *Ann. NY Acad. Sci.* **652**, 156–165 (1992).
- Ota, Y., Ando, H., Ueda, H. & Urano, A. Differences in seasonal expression of neurohypophysial hormone genes in ordinary and precocious male masu salmon. *Gen. Comp. Endocrinol.* **116**, 40–48 (1999).
- Young, L. J., Wang, Z. X. & Insel, T. R. Neuroendocrine bases of monogamy. *Trends Neurosci.* **21**, 71–75 (1998).
- Emerson, S. B. & Boyd, S. K. Mating vocalizations of female frogs: control and evolutionary mechanisms. *Brain Behav. Evol.* **53**, 187–197 (1999).
- Goodson, J. L. Vasotocin and vasoactive intestinal polypeptide modulate aggression in a territorial songbird, the violet-eared waxbill (Estrildidae: *Uraeginthus granatina*). *Gen. Comp. Endocrinol.* **111**, 233–244 (1998).
- Young, L. J., Nilsen, R., Waymire, K. G., MacGregor, G. R. & Insel, T. R. Increased affiliative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. *Nature* **400**, 766–768 (1999).
- Brantley, R. K. & Bass, A. H. Alternative male spawning tactics and acoustic signals in the plainfin midshipman *Porichthys notatus* Girard (Teleostei, Batrachoididae). *Ethology* **96**, 213–232 (1994).
- Bass, A. H. Shaping brain sexuality. *Am. Sci.* **84**, 352–363 (1996).
- Moore, F. L. & Lowry, C. A. Comparative neuroanatomy of vasotocin and vasopressin in amphibians and other vertebrates. *Comp. Biochem. Physiol. C* **119**, 251–260 (1998).
- Fine, M. L. & Perini, M. A. Sound production evoked by electrical stimulation of the forebrain in the oyster toadfish. *J. Comp. Physiol. A* **174**, 173–185 (1994).
- Grober, M. S. & Sunobe, T. Serial adult sex change involves rapid and reversible changes in forebrain neurochemistry. *NeuroReport* **7**, 2945–2949 (1996).
- Foran, C. M. & Bass, A. H. Preoptic AVT immunoreactive neurons of a teleost fish with alternative reproductive tactics. *Gen. Comp. Endocrinol.* **111**, 271–282 (1998).
- Bass, A. & Baker, R. Sexual dimorphisms in the vocal control system of a teleost fish: morphology of physiologically identified neurons. *J. Neurobiol.* **21**, 1155–1168 (1990).
- Bass, A. H., Marchaterre, M. A. & Baker, R. Vocal-acoustic pathways in a teleost fish. *J. Neurosci.* **14**, 4025–4039 (1994).
- Bass, A. H. & Baker, R. Phenotypic specification of hindbrain rhombomeres and the origins of rhythmic circuits in vertebrates. *Brain Behav. Evol.* **50**, 3–16 (1997).
- Batten, T. F. C., Cambre, M. L., Moons, L. & Vandesande, F. Comparative distribution of neuropeptide-immunoreactive systems in the brain of the green molly *Poecilia latipinna*. *J. Comp. Neurol.* **302**, 893–919 (1990).
- Goodson, J. L. Territorial aggression and dawn song are modulated by septal vasotocin and vasoactive intestinal polypeptide in male field sparrows (*Spizella pusilla*). *Horm. Behav.* **34**, 67–77 (1998).
- Moore, F. L. & Miller, L. J. Arginine vasotocin induces sexual behavior of newts by acting on cells in the brain. *Peptides* **4**, 97–102 (1983).
- Floody, O. R., Cooper, T. T. & Albers, H. E. Injection of oxytocin into the medial preoptic-anterior hypothalamus increases ultrasound production by female hamsters. *Peptides* **19**, 833–839 (1998).
- Rose, J. D., Kinnaird, J. R. & Moore, F. L. Neurophysiological effects of vasotocin and corticosterone on medullary neurons: implications for hormonal control of amphibian courtship behavior. *Neuroendocrinology* **62**, 406–417 (1995).
- Manning, M. et al. Solid-phase synthesis of 16 potent (selective and nonselective) *in vivo* antagonists of oxytocin. *J. Med. Chem.* **32**, 382–391 (1989).
- Mahlmann, S. et al. Structure, function, and phylogeny of [arg8] vasotocin receptors from teleost fish and toad. *Proc. Natl Acad. Sci. USA* **91**, 1342–1345 (1994).
- Northcutt, R. G. & Bradford, M. R. Jr in *Comparative Neurology of the Telencephalon* (ed. Ebessson, S. O. E.) (Plenum, New York, 1980).
- Hausmann, H., Meyerhof, W., Zwiers, H., Lederis, K. & Richter, D. Teleost isotocin receptor: structure, functional expression, mRNA distribution and phylogeny. *FEBS Lett.* **370**, 227–230 (1995).
- Bluthe, R. M. & Dantzer, R. Social recognition does not involve vasopressinergic neurotransmission in female rats. *Brain Res.* **535**, 301–304 (1990).
- Marler, C. A., Boyd, S. K. & Wilczynski, W. Forebrain arginine vasotocin correlates of alternative mating strategies in cricket frogs. *Horm. Behav.* **36**, 53–61 (1999).
- Swaab, D. F. & Hofman, M. A. An enlarged suprachiasmatic nucleus in homosexual men. *Brain Res.* **537**, 141–148 (1990).
- Everts, H. G. J., DeRuiter, A. J. H. & Koolhaas, J. M. Differential lateral septal vasopressin in wild-type rats: correlation with aggression. *Horm. Behav.* **31**, 136–144 (1997).
- Delville, Y., Melloni, R. H. & Ferris, C. F. Behavioral and neurobiological consequences of social subjugation during puberty in golden hamsters. *J. Neurosci.* **18**, 2667–2672 (1998).
- Winslow, J. T. & Insel, T. R. Effects of central vasopressin administration to infant rats. *Eur. J. Pharmacol.* **233**, 101–107 (1993).

Correspondence and requests for materials should be addressed to J. L. G. (e-mail: jlg14@cornell.edu).

Acknowledgements

We thank M. Weeg, M. Marchaterre and R. Wytenbach for technical assistance; J. McKibben for field support; E. Adkins-Regan and K. Reeve for comments on the manuscript; and R. Darlington for consultation on statistics. Logistical support provided by the University of California Bodega Marine Laboratory. Funding was provided by the NSF (to A.H.B.) and the NIH (postdoctoral fellowship to J.L.G.).

Mapping the conformational wave of acetylcholine receptor channel gating

Claudio Grosman, Ming Zhou & Anthony Auerbach

Department of Physiology and Biophysics, State University of New York at Buffalo, 124 Sherman Hall, Buffalo, New York, 14214, USA

Allosteric transitions allow fast regulation of protein function in living systems. Even though the end points of such conformational changes are known for many proteins, the characteristics of the paths connecting these states remain largely unexplored. Rate-equilibrium linear free-energy relationships (LFERs) provide information about such pathways by relating changes in the free energy of the transition state to those of the ground states upon systematic perturbation of the system¹. Here we present an LFER analysis of the gating reaction pathway of the muscle acetylcholine receptor. We studied the closed $\rightleftharpoons$ open conformational change at the single-molecule level following perturbation by series of single-site mutations, agonists and membrane voltages. This method provided a snapshot of several regions of the receptor at the transition state in terms of their approximate positions along the reaction coordinate, on a scale from 0 (closed-like) to 1 (open-like). The resulting map reveals a spatial gradient of positional values, which suggests that the conformational change proceeds in a wave-like manner, with the low-to-high affinity change at the transmitter-binding sites preceding the complete opening of the pore.

The thermodynamics of allosteric conformational changes are well understood: an allosteric effector binds with different affinities to (at least) two intrinsic protein conformations (often active and inactive) thus biasing the conformational equilibrium towards the high-affinity form. The structures of several allosteric proteins in their active or inactive states have been solved². However, insight regarding the pathways that connect these states cannot be gained readily from spectroscopic or crystallographic techniques, as these mainly provide information about long-lived species³. Information about the transient intermediate stages of a conformational change can be obtained from equilibrium and kinetic measurements in the form of rate-equilibrium LFERs^{1,4}. LFERs have been widely

employed in physical organic chemistry to elucidate reaction mechanisms^{1,5} and, more recently and less extensively, in biochemistry to analyse enzyme catalysis^{3,6–9}, the T $\rightleftharpoons$ R conformational change of haemoglobin¹⁰ and the chaperonin GroEL¹¹, and the folding–unfolding pathways of small soluble proteins^{12,13}. They have also been used to calculate the isomerization rate constants of unliganded and mono-liganded acetylcholine receptors¹⁴.

The concept of an LFER (Fig. 1a and Methods) can be applied to single-step reactions and is conveniently expressed in equations containing variables that are directly accessible from kinetic and equilibrium measurements^{1,4}:

$$\Delta G_{\text{opening}}^{\ddagger} = \Phi \Delta G_{\text{C}=\text{O}}^{\circ} + \Delta G_0^{\ddagger} \quad (1)$$

or

$$\Delta G_{\text{closing}}^{\ddagger} = (\Phi - 1) \Delta G_{\text{C}=\text{O}}^{\circ} + \Delta G_0^{\ddagger} \quad (2)$$

where the values on the left are the activation free energies of opening and closing, and $\Delta G_{\text{C}=\text{O}}^{\circ}$ is the equilibrium free-energy change of each receptor in the perturbation series. $\Delta G_0^{\ddagger}$, a constant for each series, is the intrinsic activation barrier (that is, $\Delta G_{\text{opening}}^{\ddagger}$ and $\Delta G_{\text{closing}}^{\ddagger}$ when $\Delta G_{\text{C}=\text{O}}^{\circ} = 0$), and Φ , the slope, is an estimate of the position of the transition state along the reaction pathway.

The muscle acetylcholine receptor (AChR) (relative molecular mass (M_r) $\approx$ 290K) is a heteromeric ($\alpha_2\beta\delta\epsilon/\gamma$) ion channel that mediates synaptic transmission by switching between ion-permeable ('open') and ion-impermeable ('closed') conformations. It is the best studied member of a superfamily of pentameric ionotropic receptors that also includes neuronal nicotinic, serotonin, glycine, γ -aminobutyric acid (GABA)_A and GABA_C receptors¹⁵. All subunits are homologous and share a characteristic topology¹⁶ (Fig. 1b).

Figure 2 shows the LFER analysis of a pore-residue mutant series at the single-channel level. The serine occupying the M2 12' position of the wild-type δ -subunit (δ 268, three residues away from a conserved central 9'-leucine) was mutated to twelve different amino acids. Instead of ACh, which opens these mutants at rates that are too high to be reliably estimated, we used two weaker agonists, choline and acetylthiocholine. We estimated rate and equilibrium constants at saturating concentrations to isolate channel isomerization from ligand-binding steps. When probed at this site, Φ is $\sim$ 0.3, which suggests that the structure of the gating transition state near this residue is a hybrid that is $\sim$ 30% like that of

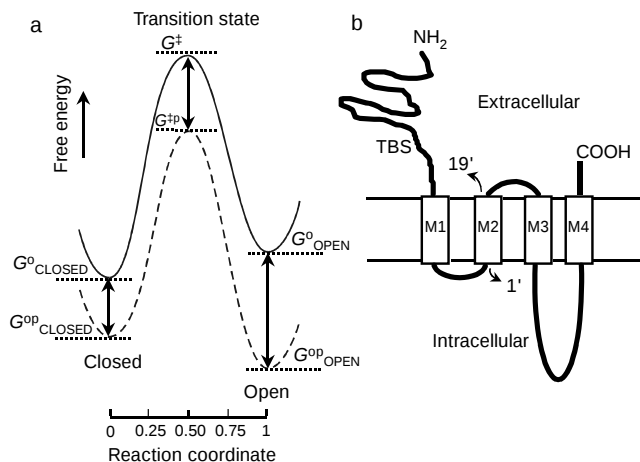


Figure 1 Schematic representations of the LFER concept and AChR topology. **a**, A hypothetical slice of the free-energy landscape of AChR gating showing the most likely (the lowest free-energy) pathway. The energy profile can be considered to be smooth and dominated by a single transition state (that is, minimally frustrated) based on the single-exponential kinetics of gating (Fig. 2b). The solid and dashed lines represent the AChR

before and after a stabilizing local perturbation (denoted by 'p'), respectively. See Methods for a fuller explanation. **b**, Current view of the ligand-gated ion-channel subunit topology. Two transmitter-binding sites (TBS) per pentamer are formed in the extracellular N-terminal domain. The channel pore is lined, for the most part, by the five M2 segments.

the open state and $\sim 70\%$ like that of the closed state. As shown in Fig. 2, this value changes little with the chemical nature of the agonist or the applied voltage. As the behaviour of the proline mutant was anomalous, it was excluded from the LFER analysis. The dotted line connects this mutant to the wild type ($\Phi = 0.874$) and is an example of the risk of interpreting LFERs generated with only two points. The mutational approach was extended to other regions of the protein and the results are summarized in Table 1.

Site-directed mutagenesis provides a picture of the transition state with resolution at the single amino-acid residue level. Other independent variables, however, can be used to probe the transition state of functionally defined domains. Different agonists, for example, can be used to explore the transition state of the transmitter-binding sites (TBS). As shown in Fig. 3, $\Phi = 0.931 \pm 0.035$ (mean $\pm$ s.d., 13 points, $K_{\max}/K_{\min} \approx 930$, $r = 0.401$) when agonists are used as perturbants. We conclude that, at the gating transition state, the TBSs have almost completely adopted the high-affinity conformation.

Another variable that can be used to explore a functional domain is the membrane potential. The AChR gating equilibrium constant increases with membrane hyperpolarization because (as yet unidentified) charged or dipolar structural elements move in the electric field during gating¹⁷. When different voltages were applied to diliganded AChRs, Φ was 0.070 ± 0.060 (mean $\pm$ s.d., 4 points, $K_{\max}/K_{\min} \approx 3$)¹⁸. This indicates that these elements are essentially in the closed configuration at the transition state of diliganded gating. When the same measurements are made in the absence of ligand,

$\Phi = 1.025 \pm 0.053$ (mean $\pm$ s.d., 8 points, $K_{\max}/K_{\min} \approx 18$) (Fig. 4) even though the effective gating charge of un- and diliganded AChR gating is identical. Thus, there seems to be a fundamental difference in the gating pathways of unliganded and diliganded AChRs; their end points, in contrast, are probably the same¹⁹.

From this limited sample, there appears to be a spatial gradient of Φ -values for the diliganded receptor, from ~ 1 at the TBS to ~ 0 within the membrane. Our interpretation of this gradient is that the opening transition of diliganded AChRs occurs as a conformational 'wave' that starts at the TBS, progresses through the transmembrane domains and ends near the structures that define voltage dependence (with the reverse wave taking place during closing). Propagated conformational rearrangements have also been described in single-step protein-folding reactions (for example, ref. 20) and near the binding site of cyclic-nucleotide-gated ion channels during gating²¹. To our knowledge, there is no *a priori* reason to have expected this particular ordering of events during AChR gating (that is, initiation at the TBS and propagation to the membrane interior) because agonist is not needed for the channel to open²². Also, there seems to be no reason why Φ could not have been the same everywhere in the protein, as would be the case if the conformational transition were synchronous. Indeed, this may occur in unliganded AChRs.

A single-pathway mechanism for gating predicts a linear rate-equilibrium plot. In contrast, a mechanism with multiple parallel pathways predicts a curved function, the curvature of which increases as the difference between the Φ -values of the individual trajectories increases. The linearity of the rate-equilibrium plots in Fig. 2 indicates that, at least as far as the $\delta 12'$ position is concerned, the closed and open forms of diliganded AChRs are connected either by a single reaction pathway or by a set of parallel pathways of

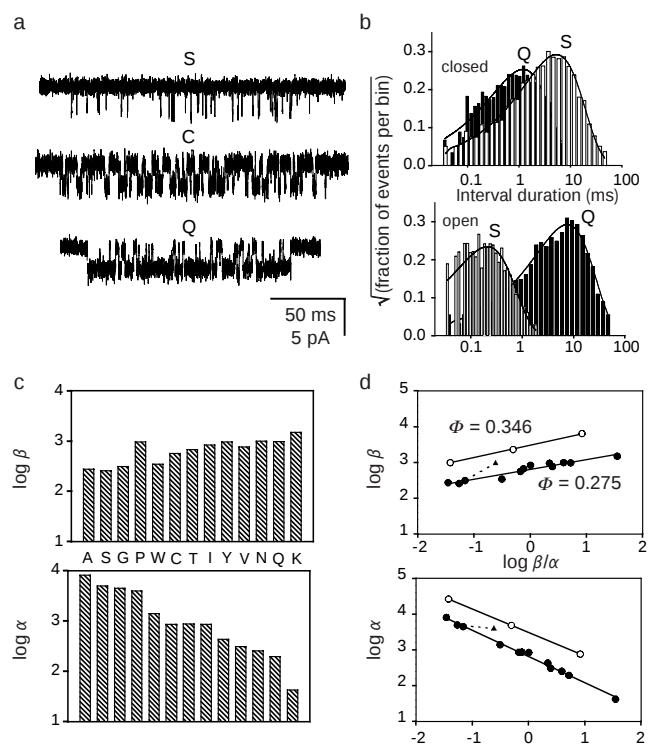


Figure 2 Site-directed mutations as a probe of the gating pathway. **a**, Representative single-channel currents of the wild-type and two example mutants (denoted by single-letter amino-acid code) activated with saturating, 20 mM choline at a membrane potential of ~ -100 mV. Openings are downwards deflections. **b**, Monoexponential dwell-time histograms and computed density functions. **c**, Logarithms of the opening (β) and closing (α) rate constants for the mutant series in increasing order of equilibrium constant ($K = \beta/\alpha$). The increase in β/α is mainly due to a decrease in α . **d**, Log k versus log β/α plots. Filled circles, the set of mutants in **c** (except $\delta S268P$, triangle) activated with 20 mM choline at ~ -100 mV; open circles, $\delta 268$ S, T and Q receptors activated with saturating, 2 mM acetylthiocholine at $\sim +50$ mV.

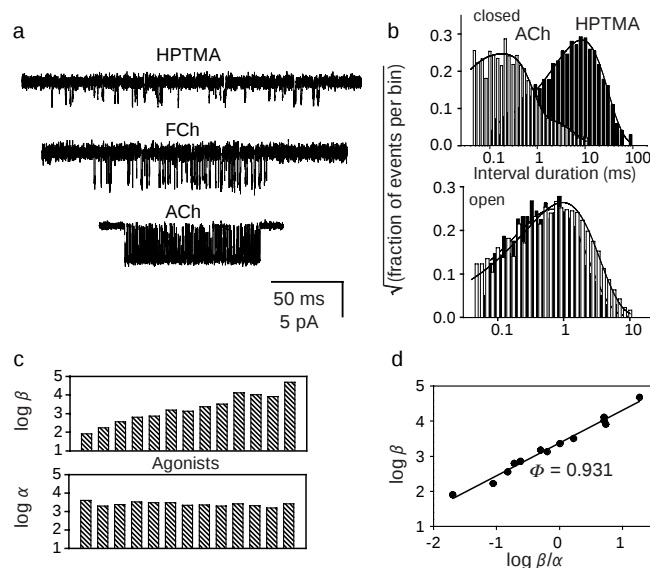


Figure 3 Agonists as a probe of the gating pathway. **a**, Representative single-channel currents of wild-type AChR activated with (3-hydroxy-propyl)-trimethylammonium (HPTMA, 10 mM, -50 mV), formylcholine (FCh, 2 mM, -50 mV) and acetylcholine (ACh, 100 μ M, -100 mV). Openings are downwards. **b**, Dwell-time histograms and computed density functions. β for ACh is ~ 350 times larger than that for HPTMA but the values of α are similar. The small, slow component of closed times for ACh does not arise from the gating reaction²⁵. **c**, Logarithms of β and α for 13 different agonists in increasing order of gating equilibrium constant. From left to right: choline, HPTMA, (2-chloro-ethyl)-trimethylammonium, (4-hydroxy-butyl)-trimethyl-ammonium, propyl-trimethylammonium, ethyl-trimethylammonium, (4-methoxy-butyl)-trimethylammonium, acetylthiocholine, FCh, (4-keto-pentyl)-trimethylammonium, carbamylcholine, tetramethylammonium and ACh. **d**, LFER plot of the agonist series.

similar fractional Φ -values. This linearity also implies that both the position of the transition state along the reaction coordinate and the intrinsic activation barrier are unaffected by the mutations^{3,8,23}. This is remarkable insofar as the linear relationship holds for twelve different side chains over a three-order-of-magnitude variation in the gating equilibrium constant. The fact that mutational effects can be ordered in an LFER is compelling evidence that the mutations in question do not alter the structure of the protein in hopelessly complicated ways²⁴. Therefore, the systematic study of such mutant

series constitutes an excellent analytical tool for probing allosteric-transition pathways with high detail.

A large-scale application of the LFER–mutagenesis approach will allow us to refine the current model of the transition state of AChR gating. This will place strong constraints on future structure-based interpretations of the gating mechanism. Application of the LFER methodology to other allosteric proteins (including voltage- and stretch-sensitive ion channels) should reveal whether initiation at the effector site and propagation to the active site is a common theme. □

Methods

Linear free-energy relationships

The concept of an LFER is based on the idea that the transition state is intermediate between the ground states with respect to the sensitivity of its free energy to structural (for example, mutations) or environmental (for example, ligands, voltage, solvent) changes^{1,4}. In symbols,

$$(G^{\text{tp}} - G^{\text{t}}) = \Phi(G_{\text{OPEN}}^{\text{op}} - G_{\text{OPEN}}^{\text{c}}) + (1 - \Phi)(G_{\text{CLOSED}}^{\text{op}} - G_{\text{CLOSED}}^{\text{c}})$$

where the transition-state's free-energy change upon perturbation (denoted by the superscript 'p') is the linear combination of the free-energy changes in the open and closed states, and Φ ($0 \leq \Phi \leq 1$) is a measure of the extent to which the transition state resembles the open state. Rearrangement of this linear-combination equation yields equations (1) and (2), which were used to estimate Φ values. In the example of Fig. 1a, $\Phi = 0.3$; that is, the stabilization of the transition state is 30% like that of the open and 70% like that of the closed state. This suggests that, when the transition state is reached, the conformational change in the vicinity of the perturbed region is ~30% of the way to the open state. The position of the transition state along the reaction coordinate for the AChR as a whole, however, is approximately given by the ensemble average of Φ values of all individual residues; in Fig. 1a, it was arbitrarily placed at ~0.5.

Single-channel electrophysiology

We recorded AChR single-channel currents in the cell-attached patch configuration²⁵ from transiently-transfected HEK-293 cells²⁶. Site-directed mutations were engineered on the α , β , δ or ϵ subunit following standard techniques. We estimated gating rate constants in the presence of saturating concentrations of choline or acetylthiocholine (Fig. 2 and Table 1) by applying a full-likelihood algorithm²⁷ that includes a correction for missed events (program MIL) to a C $\rightleftharpoons$ O model ($f_c = 18$ kHz; dead time ≈ 30 μ s). Clusters were defined as series of openings separated by closures shorter than a critical time (τ_c). For every patch, different τ_c -values were tested and the idealized intracluster dwell-time series (program SKM) were analysed using MIL. In each case, the longest τ_c defining clusters that were still best modelled by C $\rightleftharpoons$ O schemes was chosen. Fast blockade by 20 mM choline at ~ -100 mV reduced the current amplitude by $\sim 50\%$ and lengthened the openings less than two-fold. As no differences were found between the reduction in current amplitude in wild-type and mutant receptors, the prolongation of openings is also

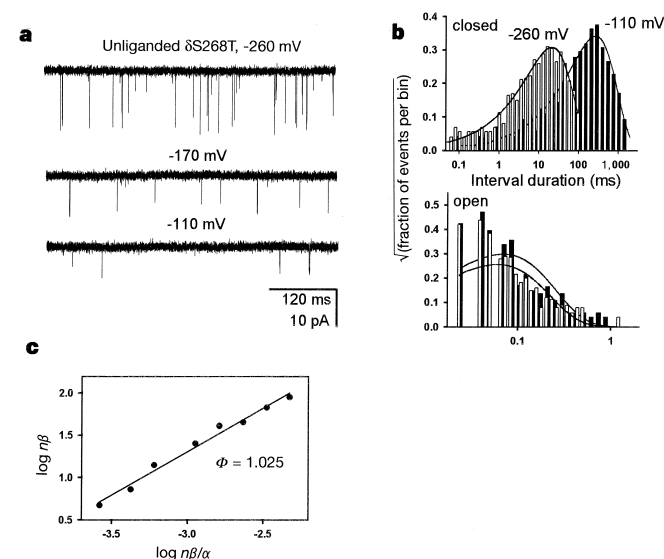


Figure 4 Voltage as a probe of the gating pathway in unliganded receptors.

a, Representative single-channel currents recorded in the absence of agonist at different voltages. The δ S268T mutation was engineered to increase spontaneous activity. In this patch the kinetics of gating were best fit by a single-step C $\rightleftharpoons$ O scheme. Openings are downwards. **b**, Dwell-time histograms and computed density functions. As the number of channels in the patch (n) was not known, the opening rate was interpreted as $n\beta$. **c**, LFER plot of the voltage series. Each point corresponds to a different membrane potential. From left to right: -80 , -110 , -140 , -170 , -200 , -210 , -230 and -260 mV. The increase in the $n\beta/\alpha$ ratio is entirely due to an increase in β .

Table 1 Φ values of diliganded gating at different positions of $\alpha_2\beta\delta\epsilon$ AChRs

Mutated residue	Putative location	Φ (mean $\pm$ s.d.)	Agonist	Equilibrium constant range*	Number of constructs†	Correlation coefficient‡
α 93§	TBS	1.157	ACh	57	2	0.040
α 198§	TBS	1.021	ACh	2	2	
α 152	TBS	0.979 ± 0.062	ACh	75	9	
ϵ 175§	TBS	0.938	ACh	35	2	
α 200¶	TBS	0.937	ACh	153	2	
α 153#	TBS	0.920	Choline	23	2	0.999
ϵ 184§	TBS	0.847 ± 0.072	ACh	56	3	
α 217#	Amino-terminal third of M1	0.839	Choline	80	2	
α 269	M2–M3 linker	0.694 ± 0.028	Choline	101	4	0.988
β 265	M2, 12' ^{‡‡}	0.663	Choline	5	2	
ϵ 269#	M2, 17'	0.591	Choline	50	2	
ϵ 278	M2–M3 linker	0.568	Choline	6	2	
β 279	M2–M3 linker	0.431	Choline	23	2	
α 285**	Middle of M3	0.418 ± 0.039	ACh	368	5	0.963
δ 268	M2, 12'	0.275 ± 0.023	Choline	1,016	12	0.945
β 266#	M2, 13'	0.266	Choline	7	2	

* $K_{\text{max}}/K_{\text{min}}$, gating equilibrium constants (K) were calculated as the ratio between the opening (β) and the closing (α) rate constants.

† Including the wild type.

‡ Correlation coefficients are between $\log \beta$ and $\log \alpha$ (not between $\log \beta$ and $\log \beta/\alpha$). These are small when Φ is close to 0 or 1 because one rate constant changes much more than the other upon perturbation. In these cases, the existence of an LFER cannot be ascertained but the interpretation of Φ as a positional value remains valid.

§ From ref. 28.

|| Transmitter-binding sites.

¶ From ref. 29.

From ref. 30.

‡‡ M2 residues are numbered 1'–19' from the intracellular to the extracellular end.

** From ref. 31.

expected to be the same for the entire perturbation series and, therefore, it does not affect the estimate of the LFER slope. Fast blockade by acetylthiocholine (more pronounced than that of choline) was relieved by depolarizing the patch to $\sim +50$ mV. For the agonist series of Fig. 3, clusters of single-channel activity were defined over a range of ligand concentrations, β was estimated from the saturating-concentration limit of the opening rate, and α was estimated from kinetic modelling, as described²⁵. Concentrations up to 2 mM were used for ACh and carbamylcholine, up to 5 mM for tetramethylammonium and (4-keto-pentyl) trimethylammonium, up to 10 mM for formylcholine (FCh) and acetylthiocholine and up to 20 mM for all other agonists. In the case of unliganded δ S268T AChRs (Fig. 4), where openings did not occur in clusters, continuous stretches of data were idealized (program SKM) and gating rate constants were estimated using the program MIL ($f_c = 18$ kHz; dead time ≈ 20 μ s). The programs are available from www.qub.buffalo.edu.

To avoid the uncertainty associated with the calculation of activation-barrier heights from rate constants, we used $\log k$ versus $\log K$ plots instead of $\Delta G^\ddagger$ versus ΔG° plots. Φ values were calculated from the slope of linear fits to such plots ('Brønsted' plots).

Received 6 September; accepted 26 November 1999.

1. Lefler, J. E. & Grunwald, E. *Rates and Equilibria of Organic Reactions* (Wiley, New York, 1963).
2. Kyte, J. *Mechanism in Protein Chemistry* 461–519 (Garland, New York, 1995).
3. Fersht, A. R., Leatherbarrow, R. J. & Wells, T. N. C. Quantitative analysis of structure–activity relationships in engineered proteins by linear free-energy relationships. *Nature* **322**, 284–286 (1986).
4. Lefler, J. E. Parameters for the description of transition states. *Science* **117**, 340–341 (1953).
5. Jaffé, H. H. A reexamination of the Hammett equation. *Chem. Rev.* **53**, 191–261 (1953).
6. Kirsch, J. F. In *Advances in Linear Free Energy Relationships* (eds Chapman, N. B. & Shorter, J.) 369–400 (Plenum, London, 1972).
7. Estell, D. A. *et al.* Probing steric and hydrophobic effects on enzyme–substrate interactions by protein engineering. *Science* **233**, 659–663 (1986).
8. Fersht, A. R., Leatherbarrow, R. J. & Wells, T. N. Structure–activity relationships in engineered proteins: analysis of use of binding energy by linear free energy relationships. *Biochemistry* **26**, 6030–6038 (1987).
9. Toney, M. D. & Kirsch, J. F. Direct Brønsted analysis of the restoration of activity to a mutant enzyme by exogenous amines. *Science* **243**, 1485–1488 (1989).
10. Eaton, W. A., Henry, E. R. & Hofrichter, J. Application of linear free energy relations to protein conformational changes: the quaternary structural change of hemoglobin. *Proc. Natl Acad. Sci. USA* **88**, 4472–4475 (1991).
11. Yifrach, O. & Horowitz, A. Mapping the transition state of the allosteric pathway of GroEL by protein engineering. *J. Am. Chem. Soc.* **120**, 13262–13263 (1998).
12. Fersht, A. R. Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.* **5**, 79–84 (1995).
13. Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z. & Wolynes, P. G. Protein folding funnels: the nature of the transition state ensemble. *Folding Design* **1**, 441–450 (1996).
14. Edelstein, S. J., Schaad, O., Henry, E., Bertrand, D. & Changeux, J. P. A kinetic mechanism for nicotinic acetylcholine receptors based on multiple allosteric transitions. *Biol. Cybern.* **75**, 361–379 (1996).
15. Ortel, M. O. & Lunt, G. G. Evolutionary history of the ligand-gated ion-channel superfamily of receptors. *Trends Neurosci.* **18**, 121–127 (1995).
16. Karlin, A. & Akabas, M. H. Toward a structural basis for the function of nicotinic acetylcholine receptors and their cousins. *Neuron* **15**, 1231–1244 (1995).
17. Magleby, K. L. & Stevens, C. F. The effect of voltage on the time course of end-plate currents. *J. Physiol.* **223**, 151–171 (1972).
18. Auerbach, A., Sigurdson, W., Chen, J. & Akk, G. Voltage dependence of mouse acetylcholine receptor gating: different charge movements in di-, mono- and unliganded receptors. *J. Physiol.* **494**, 155–170 (1996).
19. Monod, J., Wyman, J. & Changeux, J. P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
20. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for nucleation–condensation mechanism for protein folding. *J. Mol. Biol.* **254**, 260–288 (1995).
21. Sunderman, E. R. & Zagotta, W. N. Sequence of events underlying the allosteric transition of rod cyclic nucleotide-gated channels. *J. Gen. Physiol.* **113**, 621–640 (1999).
22. Jackson, M. B. Spontaneous openings of the acetylcholine receptor channel. *Proc. Natl Acad. Sci. USA* **81**, 3901–3904 (1984).
23. Straub, J. E. & Karplus, M. The interpretation of site-directed mutagenesis experiments by linear free energy relations. *Protein Eng.* **3**, 673–675 (1990).
24. Hammett, L. P. In *Advances in Linear Free Energy Relationships* (eds Chapman, N. B. & Shorter, J.) vii (Plenum, New York, 1972).
25. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch.* **391**, 85–100 (1981).
26. Salamone, F. N., Zhou, M. & Auerbach, A. A re-examination of adult mouse nicotinic acetylcholine receptor channel activation kinetics. *J. Physiol.* **516**, 315–330 (1999).
27. Qin, F., Auerbach, A. & Sachs, F. Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.* **70**, 264–280 (1996).
28. Akk, G., Zhou, M. & Auerbach, A. A mutational analysis of the acetylcholine receptor channel transmitter binding site. *Biophys. J.* **76**, 207–218 (1999).
29. Akk, G., Sine, S. & Auerbach, A. Binding sites contribute unequally to the gating of mouse nicotinic α D200N acetylcholine receptors. *J. Physiol.* **496**, 185–196 (1996).
30. Zhou, M., Engel, A. G. & Auerbach, A. Serum choline activates mutant acetylcholine receptors that cause slow channel congenital myasthenic syndromes. *Proc. Natl Acad. Sci. USA* **96**, 10466–10471 (1999).
31. Wang, H.-L. *et al.* Acetylcholine receptor M3 domain: stereochemical and volume contributions to channel gating. *Nature Neurosci.* **2**, 226–233 (1999).

Acknowledgements

This work was supported by grants from the NIH (A.A.) and the Myasthenia Gravis Foundation of America Inc. (C.G. and A.A.). We thank J. Richard, F. Sachs and S. Sine for discussions and critical comments, and F. Salamone and K. Lau for technical assistance.

Correspondence and requests for materials should be addressed to C.G. (e-mail: grosman@buffalo.edu). or A.A. (e-mail: auebach@buffalo.edu).

Positional cloning of zebrafish *ferroportin1* identifies a conserved vertebrate iron exporter

Adriana Donovan*, Alison Brownlie*, Yi Zhou*, Jennifer Shepard*, Stephen J. Pratt†, John Moynihan*, Barry H. Paw*, Anna Drejer*, Bruce Barut*, Agustín Zapata§, Terence C. Lawl, Carlo Brugnara||, Samuel E. Lux*, Geraldine S. Pinkus*, Jack L. Pinkus*, Paul D. Kingsley#, James Palis#, Mark D. Fleming*, Nancy C. Andrews*† & Leonard I. Zon*†

*Department of Medicine, Division of Hematology/Oncology, Children's Hospital and Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA

†Howard Hughes Medical Institute, Children's Hospital, Boston, Massachusetts 02115, USA

‡Washington University School of Medicine, St. Louis, MO 63110, USA

§Department of Cell Biology, Faculty of Biology, Complutense University, 28040 Madrid, Spain

||Department of Laboratory Medicine, Children's Hospital, Boston, Massachusetts 02115, USA

¶Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA

#Department of Pediatrics and Cancer Center, University of Rochester Medical Center, Rochester, New York 14642, USA

Defects in iron absorption and utilization lead to iron deficiency and overload disorders. Adult mammals absorb iron through the duodenum, whereas embryos obtain iron through placental transport. Iron uptake from the intestinal lumen through the apical surface of polarized duodenal enterocytes is mediated by the divalent metal transporter, DMT1 (refs 1–3). A second transporter has been postulated to export iron across the basolateral surface to the circulation. Here we have used positional cloning to identify the gene responsible for the hypochromic anaemia of the zebrafish mutant *weissherbst*. The gene, *ferroportin1*, encodes a multiple-transmembrane domain protein, expressed in the yolk sac, that is a candidate for the elusive iron exporter. Zebrafish *ferroportin1* is required for the transport of iron from maternally derived yolk stores to the circulation and functions as an iron exporter when expressed in *Xenopus* oocytes. Human *Ferroportin1* is found at the basal surface of placental syncytiotrophoblasts, suggesting that it also transports iron from mother to embryo. Mammalian *Ferroportin1* is expressed at the basolateral surface of duodenal enterocytes and could export cellular iron into the circulation. We propose that *Ferroportin1* function may be perturbed in mammalian disorders of iron deficiency or overload.

Two independent autosomal recessive mutations of the zebrafish hypochromic blood mutant *weissherbst* (*weh*), *weh*^{th238} and *weh*^{tp85c} were isolated as part of a large-scale screen for ethyl nitrosourea (ENU)-induced mutations that disrupt embryonic development in zebrafish⁴. These *weh* mutants were identified in a morphologic screen for defects in circulating erythroid cells⁵. The number of circulating erythroid cells of both mutant alleles are normal at 33

and 48 hours post fertilization (h.p.f.), but the *weh* mutant cells are hypochromic (lacking red colour) (data not shown). Mutant embryos have little, if any, haemoglobin compared with wild-type embryos at 33 h.p.f. and 48 h.p.f. (Fig. 1a, *o*-dianisidine stain), but some haemoglobin is detectable at 72 h.p.f. The *weh*^{th238} allele has less *o*-dianisidine staining compared with the *weh*^{tp85c} allele at 72 h.p.f. (data not shown), indicating that it may be a more severe allele. In addition to the hypochromia, a progressive decrease in red cell number occurs after 48 h.p.f. By 96 h.p.f., the number of circulating erythroid cells in mutants decreases to ~20% of the number in the wild type (data not shown). The *weh* mutant cells at days 2 and 3 have a large nucleus and basophilic cytoplasm characteristic of more immature erythroid cells, and on day 5 the remaining mutant cells are misshapen (Fig. 1b). Further studies of erythroid differentiation reveal that embryonic globin messenger RNA levels are abnormally maintained during maturation (Fig. 1a). Although *weh* mutants have no gross organ defects in addition to their anaemia, all mutant embryos die between day 7 and day 14 of development (data not shown).

Given the possibility that hypochromia could result from iron deficiency, we measured iron levels in *weh* erythroid cells. The level of iron measured in 10⁴ wild-type cells ranged from 0.093 ng to 0.208 ng (*n* = 3), whereas the levels of iron in the same number of *weh* mutant cells ranged from 0.014 ng to 0.033 ng (*n* = 3). The 4–9-fold decrease in erythrocyte iron levels could be due to low levels of iron in circulation. As confirmation of this hypothesis, iron–

dextran injected intravenously into *weh*^{tp85c} mutant embryos completely rescued haemoglobin production (Fig. 1c). This rescue shows that *weh* mutant erythroid cells are fully capable of haemoglobinization, and that the hypochromia is caused by inadequate circulatory iron levels.

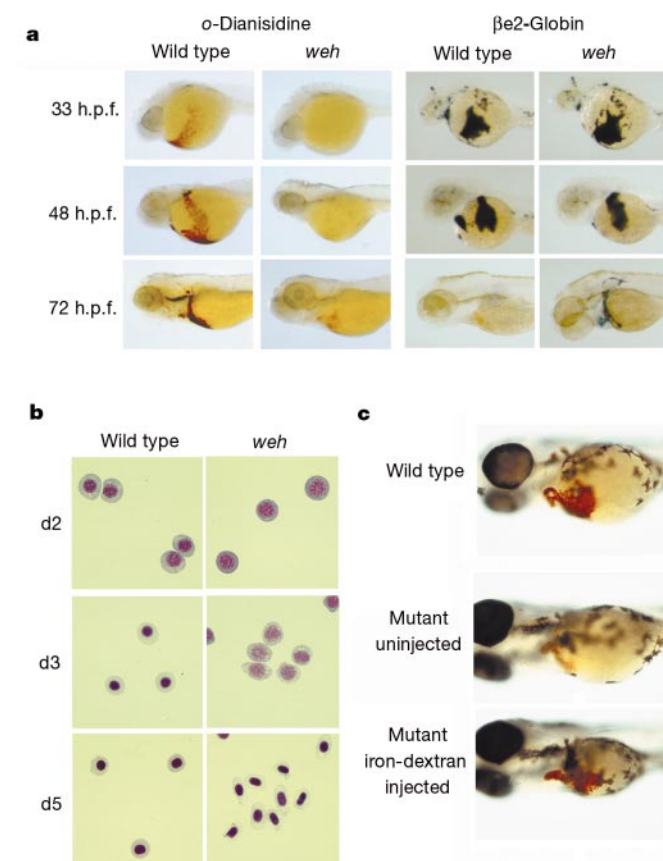


Figure 1 Phenotypic analysis of *weisherbst* (*weh*) mutants. **a**, Detection of haemoglobin (*o*-dianisidine staining) and β 2-globin mRNA (whole embryo *in situ* hybridization) in wild-type and mutant embryos at 33 hours post fertilization (h.p.f.), 48 h.p.f. and 72 h.p.f. **b**, Wright–Giemsa staining of circulating embryonic erythroid cells from wild-type and *weh*^{tp85c} embryos on day 2 (56 h.p.f.), day 3 (80 h.p.f.) and day 5 (125 h.p.f.). **c**, Injection of iron–dextran into *weh*^{tp85c} mutants. Embryos were stained for haemoglobin (*o*-dianisidine) at 80 h.p.f.

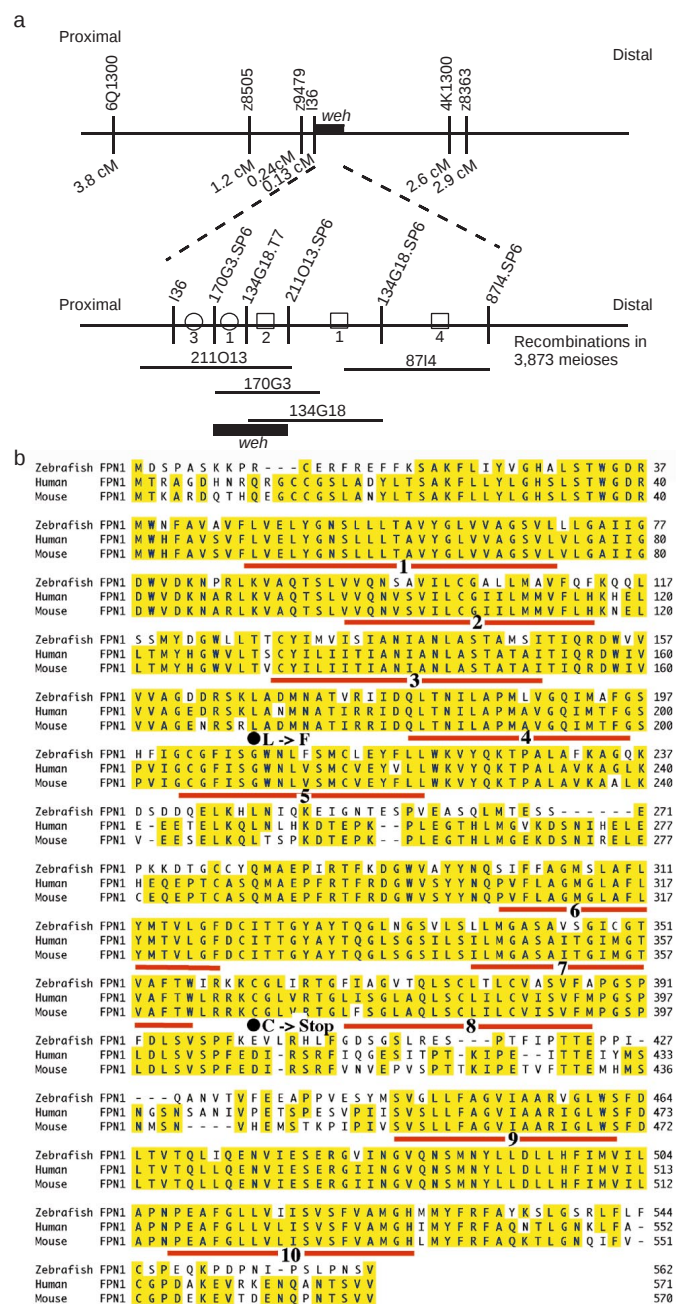


Figure 2 Positional cloning of the *weisherbst* gene. **a**, The *weh* locus is depicted by a thick black bar just distal to the AFLP marker I36. Below is an enlarged view of the *weh* locus that depicts the BAC and PAC genomic clones identified by a chromosomal walk in an analysis of 3,873 meioses. Genotyping of a total of 1,783 meioses from haploid animals and 2,090 meioses from diploid mutants narrowed the critical interval containing the gene to the PAC clones 211013 and 170G3. The numbers of recombination events identified on the proximal side (circles) and distal side (squares) of the *weh* locus are indicated. **b**, Sequence alignment of zebrafish, human and mouse Ferroporin1 (FPN1). The initiator methionine in all three species was established by the presence of upstream, in-frame stop codons. Identical amino acids (yellow), predicted transmembrane domains (red bars below sequence), and mutations identified in the *weh*^{th238} and *weh*^{th85c} alleles (black circles below sequence) are indicated.

To gain further insight into this phenotype, we isolated the *weh* mutant gene by positional cloning methods. Study of the segregation of centromeric microsatellite markers in half-tetrad gynogenetic diploid embryos localized *weh* to linkage group 9 (data not shown). Genetic mapping placed the *weh* locus in an ~6 cM interval between the random amplified polymorphic DNA (RAPD) markers 4K1300 and 6Q1300 (Fig. 2a). We isolated more closely linked markers using amplified fragment length polymorphism (AFLP) analysis (data not shown)^{6,7}, scanning ~10,000 polymorphic loci. Single-strand conformational polymorphism analysis showed that the AFLP marker I36 was 0.13 cM proximal to the *weh* locus (Fig. 2a).

A chromosomal walk towards the gene was initiated from the I36 marker (Fig. 2a), which resulted in the identification of a critical interval that contained the *weh* gene (Fig. 2a). In an attempt to identify potential candidates for the *weh* gene, we used a hybridization strategy to screen complementary DNA libraries for genes located on the PAC clones identified in the region of the *weh* locus. We hybridized a radiolabelled insert of PAC clone 170G3 to zebrafish gridded cDNA libraries. Screening of 100,000 gridded cDNA clones identified five clones of a new cDNA, designated *WC1* (for *weh* cDNA 1). Analysis of *WC1* mRNA expression in embryos and sequencing *WC1* from *weh*^{th238} mutants suggested that *WC1* was not a candidate for *weh* (data not shown). Two genes, *STAT1* and *glutaminase*, were isolated by hybridization of the zebrafish PAC clone 87I4 (Fig. 2a) to gridded cDNA libraries (data not shown). The human orthologues of *WC1*, *STAT1* and *glutaminase* are localized to human chromosome 2 in a 2.4 cM interval (data not shown), showing conserved chromosomal synteny among vertebrates⁸. Homology searches identified a pufferfish (*Fugu rubripes*) cosmid clone (121D21) that contained both *WC1* and *STAT1* (data not shown). This cosmid also contained *Fugu* homologues of other genes located on human chromosome 2. Using primers designed to the pufferfish sequence of one of these genes (121D21aB3), a 200-base pair (bp) fragment of the zebrafish orthologue was amplified from PAC 211013. A full-length cDNA (3.7 kb) of this gene, hereafter referred to as *ferroportin1*, was isolated from a zebrafish kidney cDNA library.

This gene, *ferroportin1*, has a predicted open reading frame of 562 amino acids (Fig. 2b). Sequence analysis of the *weh*^{th238} allele identified a C-to-A nucleotide transversion that causes premature termination of translation at codon 361 (Fig. 2b). Similar analysis of the *weh*^{tp85c} allele identified a single amino-acid change, Leu167→Phe, resulting from a G-to-T nucleotide difference (Fig. 2b). The finding of a premature stop mutation in *weh*^{th238} strongly suggests that the *weh* mutant phenotype is caused by a defect in *ferroportin1*. Mouse and human *ferroportin1* cDNA clones were obtained by polymerase chain reaction with reverse transcriptase (RT-PCR) of RNA isolated from liver and placenta, respectively. A conserved sequence, predicted to form a hairpin-loop structure typical of iron response elements⁹, was identified in the 5' untranslated region (UTR) of the cDNAs from all three species (data not shown). On the basis of protein structure-prediction analysis, Ferroportin1 contains at least 10 transmembrane segments (Fig. 2b).

In situ hybridization analysis of zebrafish embryos shows that *ferroportin1* mRNA is not expressed in erythroid cells. *Ferroportin1* mRNA is detected at 18 h.p.f. through 48 h.p.f. in the yolk syncytial layer (YSL) (Fig. 3a). The YSL is the peripheral layer of the yolk cell that lies just below the membrane¹⁰. This layer surrounds the entire yolk of the embryo and consists of yolk-free cytoplasm and nuclei. Yolk has been shown to contain nutrients needed during development¹¹, including iron^{11–13} (L.I.Z., unpublished data). Embryos express *ferroportin1* in the region of the YSL at 18 h.p.f. (Fig. 3a, arrow) that lies just below the developing haematopoietic cells in the intermediate cell mass¹⁴. At 48 h.p.f., *ferroportin1* is expressed in a localized region of the YSL (Fig. 3a, arrowhead). Expression is in the same region of the YSL over which the blood

flows¹⁵. At both time-points, *ferroportin1* is expressed in the region of the YSL adjacent to the blood, but not by the entire YSL. This pattern of expression suggests that *ferroportin1* expression and function in the YSL are linked to red blood cell development. Considering the iron–dextran rescue of haemoglobin production in *weh* mutants, the YSL expression of *ferroportin1* suggested that the gene might function in the transport of iron from the yolk to the embryonic circulation.

To determine that defects in the *ferroportin1* gene cause the *weh* mutant phenotype, we injected a plasmid that expresses Ferroportin1 fused to green fluorescent protein (GFP) into the yolk cell between the 256-cell and 1,000-cell stages¹⁰. After 48 hours of development, 33% of injected embryos expressed GFP strictly in the YSL. At 80 h.p.f., the phenotype of mutant embryos expressing GFP was compared to the uninjected mutants. The Ferroportin1–GFP-expressing mutant embryos (*n* = 9) had considerably more haemoglobin expression than uninjected mutants (Fig. 3b). This partial rescue of the hypochromia provides further evidence that *ferroportin1* is the *weh* gene, and shows that Ferroportin1 acts in the

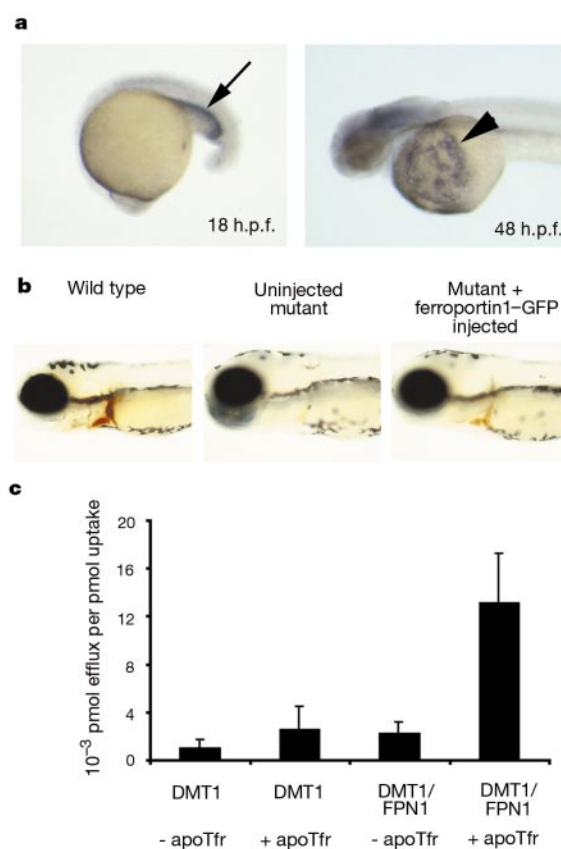


Figure 3 Developmental expression of *Ferroportin1* and phenotypic rescue of *weh* embryos. **a**, Whole embryo mRNA *in situ* hybridization analysis for *ferroportin1* at 18 and 48 h.p.f. Dark purple stain indicates expression of *ferroportin1*. Expression is seen in the YSL of embryos at both 18 and 48 h.p.f., and in the brain at 48 h.p.f. **b**, Phenotypic rescue of *weh* mutant embryos. Offspring of *weh*^{th238} heterozygote animals injected in the YSL with a plasmid expressing Ferroportin1–GFP. Embryos were stained with α -dianisidine at 80 h.p.f. to detect haemoglobin. **c**, Iron efflux measured in *Xenopus* oocytes. Oocytes expressing either DMT1 alone or DMT1 and FPN1 were loaded with ⁵⁵Fe by incubation in uptake buffer containing 60 μ M ⁵⁵FeCl₂. Efflux from individual oocytes was measured by incubation of oocytes in 500 μ l of efflux buffer with or without 20 mg ml⁻¹ apo-transferrin (– apoTfr and + apoTfr). After efflux, the total ⁵⁵Fe content of both the efflux solution and the individual oocytes was measured by scintillation counting. The data are expressed as an average (*n* = 6) of the ratio of the picomoles of efflux to the picomoles of uptake per oocyte.

YSL. The rescue of the *weh* mutant phenotype by intravenous iron-dextran injection and by Ferroportin1 expressed in the YSL indicates that Ferroportin1 functions to deliver yolk iron into the embryonic circulation.

We tested the function of Ferroportin1 using a *Xenopus* oocyte expression system. Because the proposed function of Ferroportin1 is to export iron, we needed to first load the oocytes with ^{55}Fe . Radioactive iron loading was accomplished through the expression of the iron transporter DMT1 in oocytes and then loading of ^{55}Fe at pH 5.5. ^{55}Fe -loaded oocytes that expressed either DMT1 alone or DMT1 and Ferroportin1 were tested for iron export activity either in the presence or absence of apo-transferrin, an iron chelator. To normalize for the iron content in each individual oocyte, the ratio of efflux to uptake was calculated. Our results showed that, in the presence of apo-transferrin, the efflux to uptake ratio in oocytes expressing Ferroportin1 was fivefold greater than control oocytes not expressing Ferroportin1 (Fig. 3c, $P < 0.001$).

To evaluate a potential role for *ferroportin1* in iron transport in mammals, we examined tissue expression. Northern blot analysis showed that the highest levels of expression are in human placenta, liver, spleen and kidney (Fig. 4a). In mice, *ferroportin1* mRNA is expressed specifically in the duodenum but not in the jejunum or ileum (Fig. 4b). Additionally, *ferroportin1* is expressed in the large intestine. Most intestinal iron absorption occurs in the proximal duodenum, placing *ferroportin1* in a physiologically appropriate

location to function in intestinal iron absorption. Messenger RNA *in situ* hybridization analysis was carried out on sections of mouse embryos. The primitive erythroblasts derived from the blood islands do not express *ferroportin1*, whereas the trophoblast cells of the inner placenta (Fig. 4c) express high levels of *ferroportin1* RNA. Within the embryo proper, *ferroportin1* transcripts are detected in several tissues, particularly in the gut and liver (Fig. 4d). The expression of *ferroportin1* in placenta, duodenum and liver, all prominent sites of iron transport, is consistent with the proposed role of the gene in iron transport.

To characterize the expression of Ferroportin1 protein, we generated a specific polyclonal antibody. In the human placenta, Ferroportin1 protein was primarily expressed in a basal location within the syncytiotrophoblasts (Fig. 4e). The basal surface of the syncytiotrophoblast interfaces with the fetal circulation, whereas the apical surface contacts the maternal circulation. The mammalian placenta and the zebrafish YSL provide a homologous function, serving as the site of iron transfer between mother and embryo. These results, together with the functional data in zebrafish and *Xenopus*, indicate that Ferroportin1 probably exports iron from the syncytiotrophoblast into the embryonic circulation.

A similar analysis of mouse duodenum showed Ferroportin1 staining in enterocytes in the villus (Fig. 4g). The intensity of staining was stronger at the tip of the villus compared with the crypt. Staining was particularly strong at the basolateral surface of

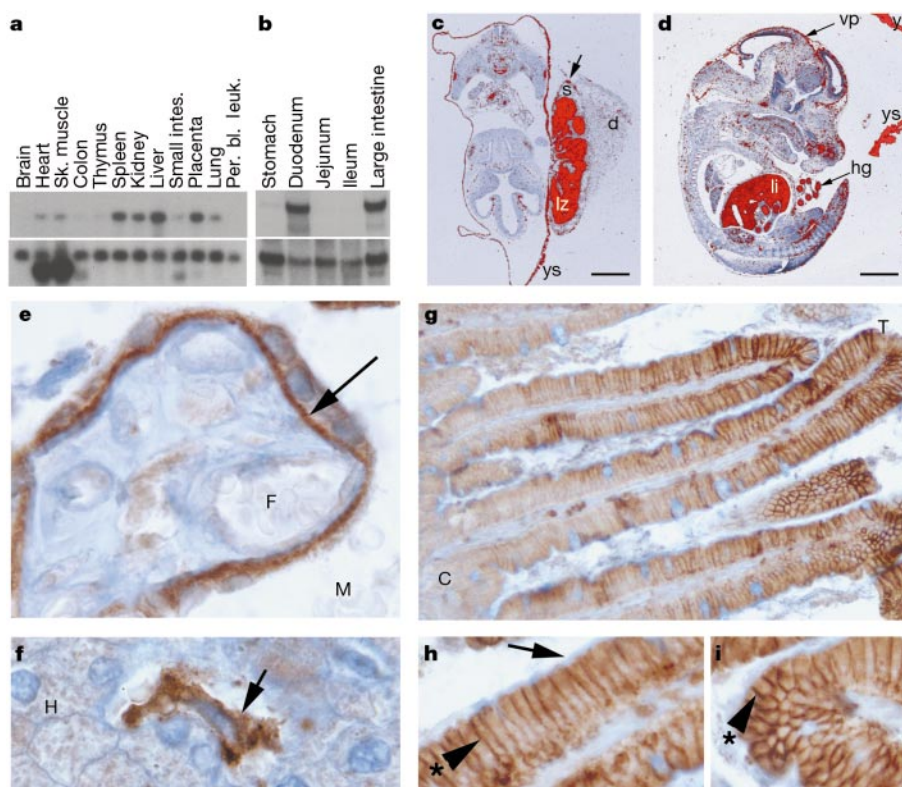


Figure 4 *Ferroportin1* expression in human and mouse tissues. **a,b**, Northern blots of adult human tissues (**a**) and mouse intestinal segments (**b**) probed with human or mouse *ferroportin1*, respectively (top panel) and mouse β -actin (lower panel). Sk, skeletal; Per. bl. leuk., peripheral blood leukocytes. **c**, mRNA *in situ* hybridization study of E12.5 mouse embryo and placenta. *Ferroportin1* transcripts are in the inner placenta (labyrinth zone, lz) and the trophoblast giant cells (arrow) at the border between the outer placenta (spongiotrophoblast, s), but not in the maternal deciduum (d). *Ferroportin1* expression is also present in the visceral endoderm of the yolk sac (ys) surrounding the embryo proper. **d**, Sagittal section of E14.5 mouse embryo. Within the embryo, *ferroportin1* expression is evident in the liver (li) and the internal and herniated gut (hg), as well as in the vascular plexus (vp) surrounding the central nervous system. Scale bars, 1 mm (**c,d**).

e-i, Immunohistochemical staining of human and mouse tissues with a Ferroportin1 peptide antibody. **e**, Human placenta stained by immunoperoxidase staining (brown staining) using an affinity purified, anti-human Ferroportin1 polyclonal rabbit antiserum (original magnification $\times 100$). The arrow indicates basal localization of staining in the syncytiotrophoblast. F, fetal circulation; M, maternal circulation. **f**, Human liver (original magnification $\times 100$). Note strong staining of Kupffer cell (arrow). Hepatocytes (H) are also positive. **g**, Mouse duodenum (original magnification $\times 40$). T, tip of villus; C, crypt. **h,i**, Higher magnification of duodenal enterocytes at the tip of the villus. Note complete lack of staining at the apical brush border (black arrow) and prominent basolateral staining (arrows with asterisk).

the enterocyte (Fig. 4h,i). The duodenal enterocytes of the small intestine are polarized epithelial cells that transport iron into the intestinal capillaries through the basolateral membrane. The mechanism of intestinal basolateral iron transport has not been established. Sex-linked anaemia (*sla*) mice have a defect in basolateral iron transport in the duodenum as deduced from ferrokinetic studies and the presence of abnormal iron deposits in duodenal enterocytes¹⁶. Analogous to *weh* mutants, the *sla* mouse has a defect in transport of maternal iron to the embryonic circulation¹⁷. The *sla* phenotype is due to a mutation in the membrane-bound multicopper ferroxidase gene, *hephaestin*¹⁸. In *Saccharomyces cerevisiae*, the hephaestin-like ferroxidase *FET3* is required for high-affinity iron uptake by the iron transporter *FTR1* (refs 19, 20). Expression of Ferroportin1 at the basolateral surface of duodenal enterocytes in mouse and the multiple-transmembrane structure of the protein make it an excellent candidate to function as a basolateral iron transporter.

Additional data from the *weh* mutant suggests that Ferroportin1 functions in the intestine of the adult zebrafish. Both *in situ* hybridization studies and immunohistochemistry showed expression of zebrafish *ferroportin1* in the intestine (data not shown). In addition, iron-dextran-rescued mutant embryos live past the normal time of lethality (day 7–14). We have successfully raised these rescued embryos to adulthood. These fish are smaller than their wild-type siblings and have a profound hypochromic anaemia (data not shown). As these fish eat a normal diet replete in iron, but nonetheless are severely anaemic, these data suggest that the gene is required for intestinal iron absorption in addition to yolk sac transport. On the basis of the basolateral expression pattern of Ferroportin1 in mammalian enterocytes and the implication that *ferroportin1* is required for intestinal iron transport in zebrafish, the protein is probably involved in iron export from enterocytes in mammals. Further experiments are required to determine whether Ferroportin1 cooperates with hephaestin in these cells.

Other tissues may use *ferroportin1* as an iron exporter. Very high levels of expression are evident in Kupffer cells (Fig. 4f), the resident macrophages of the liver and macrophages located within the splenic red pulp (data not shown). Ferroportin1 might have a role in iron export from macrophages, a critical function in recycling of iron from senescent erythrocytes.

Iron is required for many cellular processes, but it can also be toxic when present in excess; thus, iron homeostasis must be strictly maintained. We have used zebrafish genetics to identify the multiple-transmembrane domain protein Ferroportin1, a new iron export protein. In the mammalian yolk sac and placenta, Ferroportin1 may have an important conserved role in the transport of iron from the maternal to the embryonic circulation. In adults, Ferroportin1 probably functions in iron transport at the basolateral surface of duodenal enterocytes. In disorders such as iron deficiency or overload, tissues respond by altering normal iron metabolism. Ferroportin1 might be involved in the pathophysiology of iron deficiency anaemias or iron overload syndromes, such as haemochromatosis. □

Methods

Zebrafish strains and studies

Linkage analysis was performed on haploid or diploid embryos obtained from AB/DAR, AB/SJD or AB/WIK hybrids²¹. Wright–Giemsa and *o*-dianisidine staining of embryos were done as described². *In situ* hybridization analysis was done as described²².

Genetic mapping and genotyping and library screens

Linkage to centromeric markers²³ was carried out by half-tetrad analysis²⁴. For fine genetic mapping, haploids were genotyped on the proximal side of the locus with one of the RAPD markers, 4W1600, 6Q1300 or 4AC800, and on the distal side with 4K1300 or O61020 (Operon Technologies). Diploid mutant embryos were genotyped on the proximal side with the microsatellite markers z8505 or z9479 and on the distal side with z8363 (ref. 25). All library screens were done as described²⁹.

In situ hybridization and rescue experiment embryos (see below) were genotyped using allele-specific oligonucleotide (ASO) hybridization assays^{26,27} specific to the *weh*^{th238} and *weh*^{th238} *ferroportin1* alleles. The *weh*^{th238} oligonucleotides were developed to distinguish the C-to-A mutation in the *weh*^{th238} allele from the wild-type allele (wild-type 5'-AAAGAAGTGGCGGCTCATC-3' and mutant 5'-AAAGAAGTGGCGGCTCATC-3'). The *weh*^{th238} oligonucleotides were developed to distinguish the G-to-T mutation in the *weh*^{th238} allele from the wild-type allele (wild-type 5'-GAGCAAATGGCAGGTAAG-3' and mutant 5'-GAGCAAATTTGCAGGTAAG-3').

Isolation of the mouse and human *ferroportin1* cDNAs

Expressed sequence tag (EST) clones were identified that contained the 5' end (GenBank accession number D63209) and 3' end (GenBank W23461) of human *ferroportin1* and the 3' end of mouse *ferroportin1* (GenBank AA500296). The coding region of human and mouse *ferroportin1* cDNAs were cloned by RT-PCR with a forward primer made to the conserved iron response element sequence in the 5' UTR (5'-CAACTTCAGTACAGT-GTTAG-3') and a reverse primer just 3' of the stop codon of each cDNA (mouse 5'-TTATACAACAGATGTATTCGGT-3' and human 5'-AACTGTCTCAACAACAGATG-3').

Embryo injection experiments

We created a zebrafish Ferroportin1–GFP fusion protein construct by PCR. The forward PCR primer contained the start codon, 5'-CCGCTCGAGAACGCACAATGGACAGC-CCTG-3'. The reverse primer contained the last codon, 5'-CCGCTCGAGTACAGAG-TTTGGAAGTGAGGG-3'. The PCR product was subcloned into the GFP expression vector pEGFP-N1 (Clontech). Embryos from a cross of two *weh*^{th238} heterozygotes were injected²¹ with the Ferroportin1–GFP plasmid (300 ng ml⁻¹). For the iron–dextran rescue experiment, 48-h mutant embryos from a *weh*^{th238} cross were injected intravenously with an iron–dextran solution (100 mg ml⁻¹, Sigma).

Xenopus oocyte injections and ⁵⁵Fe efflux experiments

Complementary RNA for injection was prepared using the mMessage Machine kit (Ambion), using a construct containing either the rat *DMT1* cDNA in pSPORT1 (gift from H. Gunshin²) or the zebrafish *ferroportin1* cDNA in the vector pXT7 (gift from S. Sokol). Defollicularized oocytes were incubated in ND96 (96 mM NaCl, 2 mM KCl, 1.8 mM CaCl₂, 1 mM MgCl₂, 5 mM HEPES and 2.5 mM sodium pyruvate, pH 7.4) and injected with either 20 ng of *DMT1* cRNA alone or 20 ng each of both the *DMT1* and *ferroportin1* cRNAs. ⁵⁵Fe uptake and efflux were carried out 48 h after injection. ⁵⁵Fe uptake was performed in 500 µl of a solution containing 60 µM ⁵⁵FeCl₂, 100 mM NaCl, 10 mM Hepes and 1 mM ascorbic acid at pH 5.5 for 30 minutes. ⁵⁵Fe uptake was stopped by incubation of the oocytes in 1 mM cold FeCl₂, 100 mM NaCl, 10 mM HEPES and 1 mM ascorbic acid, pH 6.0, for 30 min. Individual oocytes were placed in 500 µl of either efflux buffer (100 mM NaCl, 10 mM HEPES, pH 7.4) alone or efflux buffer containing a final concentration of 20 mg ml⁻¹ apo-transferrin for 60 min. After incubation the ⁵⁵Fe levels of both the efflux solution and the individual oocytes lysed in 10% SDS were measured by scintillation counting.

Mouse *in situ* hybridization and northern blot analysis

In situ hybridization of murine embryos was performed as described²⁸. An adult multi-tissue human northern blot (Clontech) containing 2 µg of poly(A)⁺ mRNA per lane was probed with a human *ferroportin1* EST (GenBank W05488). The mouse intestinal northern blot containing 20 µg of total RNA per lane was probed with a mouse *ferroportin1* EST (GenBank AA500296).

Antibodies and immunohistochemistry

A rabbit polyclonal antibody was generated to a peptide consisting of the C-terminal 19 amino acids of the human Ferroportin1 protein (Genemed Synthesis). Antiserum was affinity purified against the peptide. Formalin or Bouin's fixed paraffin-embedded specimens were deparaffinized and heat treated for 30 min in 1.0 mM EDTA, pH 8.0, in an HS80 steamer (Black and Decker). Endogenous peroxidase activity was quenched in methanol and 3% hydrogen peroxide (5:1, v/v). The slides were then incubated in 3% normal swine serum in 0.05 M Tris, pH 7.6, followed by rabbit anti-Ferroportin1 antibody (7 µg ml⁻¹), and sequentially in HRP-conjugated goat anti-rabbit immunoglobulins (1:40 dilution, Dako), HRP-conjugated rabbit anti-goat immunoglobulins (1:40 dilution, Dako), followed by HRP-conjugated swine anti-rabbit immunoglobulins (1:52 dilution, Dako), each prepared in 0.10 M Tris, pH 7.6, containing (4%) human AB serum. Antibody localization was determined using DAB in 0.5 M Tris (pH 7.6) containing (0.035%) hydrogen peroxide. Slides were counterstained with methyl green. Control samples were incubated with normal rabbit serum or purified rabbit immunoglobulins at a protein concentration equal to the antibody preparation. In addition, pre-incubation of the antibody with specific peptide at a 550-fold molar excess neutralized the reactivity.

Red cell iron measurement

Iron levels in red blood cells were measured by atomic absorption spectrometer model 3030 equipped with Zeeman graphite furnace and autosampler (Perkin Elmer). A cell sample (at least 30,000 cells in 20 µl PBS) or blank (PBS) was diluted with 40 µl of 480 mg dl⁻¹ magnesium nitrate (matrix modifier); 20 µl of the mixture was injected into the instrument and analysed in duplicate. The instrument was calibrated with iron standards of 10, 25, 100 and 250 ng ml⁻¹ prepared from Atomic Spectroscopy Standard (Perkin Elmer).

Structure prediction

Hydropathy plots (Kyte–Doolittle) were obtained using the Genetics Computer Group (GCG) programs PEPTIDESTRUCTURE and PEPLOT with a hydropathy window of 14. Transmembrane amino-acid segments were identified and their topographies predicted using the programs PHDhtm (www.embl-heidelberg.de/predictprotein/predictprotein.html), HMMTOP (www.enzim.hu/hmmtop/), TMHMM (www.cbs.dtu.dk/services/TMHMM-1.0/), TMPred (www.ch.embnet.org/software/TMPRED_form.html), TopPred2 (www.biokemi.se/~server/toppred2/) and SOSUI (www.tuat.ac.jp/~mitaku/). Most analyses predict 10 transmembrane segments. Some predict that the first and last transmembrane segments are split and that Ferroportin contains 12 segments. Others predict that transmembrane segments 4 and 5 are not split and that only 9 segments are present.

Received 16 December 1999; accepted 19 January 2000.

- Fleming, M. D. *et al.* Microcytic anemia mice have a mutation in Nramp2, a candidate iron transporter gene. *Nature Genet.* **16**, 383–386 (1997).
- Gunshin, H. *et al.* Cloning and characterization of a mammalian proton-coupled metal-ion transporter. *Nature* **388**, 482–488 (1997).
- Andrews, N. C. Disorders of iron metabolism. *N. Engl. J. Med.* **341**, 1986–1995 (1999).
- Haffter, P. *et al.* The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**, 1–36 (1996).
- Ransom, D. G. *et al.* Characterization of zebrafish mutants with defects in embryonic hematopoiesis. *Development* **123**, 311–319 (1996).
- Ransom, D. G. & Zon, L. I. In *The Zebrafish: Genetics and Genomics* (eds Detrich, H. W. I., Westerfield, M. & Zon, L. I.) 195–210 (Academic, San Diego, 1999).
- Vos, P. *et al.* AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414 (1995).
- Postlethwait, J. H. *et al.* Vertebrate genome evolution and the zebrafish gene map. *Nature Genet.* **18**, 345–349 (1998).
- Eisenstein, R. S. & Blemings, K. P. Iron regulatory proteins, iron responsive elements and iron homeostasis. *J. Nutr.* **128**, 2295–2298 (1998).
- Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310 (1995).
- Richards, M. P. Trace mineral metabolism in the avian embryo. *Poult. Sci.* **76**, 152–164 (1997).
- Dumont, J. N. Oogenesis in *Xenopus laevis* (Daudin). VI. The route of injected tracer transport in the follicle and developing oocyte. *J. Exp. Zool.* **204**, 193–217 (1978).
- Craik, J. C. Levels of calcium and iron in the ovaries and eggs of cod *Gadus morhua* L. and plaice *Pleuronectes platessa* L. *Comp. Biochem. Physiol. A* **83**, 515–517 (1986).
- Al-Adhami, M. A. & Kunz, Y. W. Ontogenesis of haematopoietic sites in *brachydanio rerio* (Hamilton–Buchanan) (Teleostei). *Develop. Growth Differ.* **19**, 171–179 (1977).
- Rieb, J.-P. La circulation sanguine chez l'embryon de *Brachydanio rerio*. *Annales d'Embryologie et de Morphogenese* **6**, 43–54 (1973).
- Bannerman, R. M. Genetic defects of iron transport. *Fed. Proc.* **35**, 2281–2285 (1976).
- Kington, P. J., Bannerman, C. E. & Bannerman, R. M. Iron deficiency anaemia in newborn *sla* mice: a genetic defect of placental iron transport. *Br. J. Haematol.* **40**, 265–276 (1978).
- Vulpe, C. D. *et al.* Hephaestin, a ceruloplasmin homologue implicated in intestinal iron transport, is defective in the *sla* mouse. *Nature Genet.* **21**, 195–199 (1999).
- Askwith, C. *et al.* The *FET3* gene of *S. cerevisiae* encodes a multicopper oxidase required for ferrous iron uptake. *Cell* **76**, 403–410 (1994).
- Stearman, R., Yuan, D. S., Yamaguchi-Iwai, Y., Klausner, R. D. & Dancis, A. A permease–oxidase complex involved in high-affinity iron uptake in yeast. *Science* **271**, 1552–1557 (1996).
- Westerfield, M. *The Zebrafish Book* (Univ. Oregon Press, Eugene, 1993).
- Thompson, M. A. *et al.* The *cloche* and *spadetail* genes differentially affect hematopoiesis and vasculogenesis. *Dev. Biol.* **197**, 248–269 (1998).
- Knapik, E. W. *et al.* A microsatellite genetic linkage map for zebrafish. *Nature Genet.* **18**, 338–343 (1998).
- Johnson, S. L., Africa, D., Horne, S. & Postlethwait, J. H. Half-tetrad analysis in zebrafish: mapping the *ros* mutation and the centromere of linkage group I. *Genetics* **139**, 1727–1735 (1995).
- Shimoda, N. *et al.* Zebrafish genetic map with 2000 microsatellite markers. *Genomics* **58**, 219–232 (1999).
- Farr, C. J., Saiki, R. K., Erlich, H. A., McCormick, F. & Marshall, C. J. Analysis of RAS gene mutations in acute myeloid leukemia by polymerase chain reaction and oligonucleotide probes. *Proc. Natl Acad. Sci. USA* **85**, 1629–1633 (1988).
- Wood, W. L., Gitschier, J., Lasky, L. A. & Lawn, R. M. Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. *Proc. Natl Acad. Sci. USA* **82**, 1585–1588 (1985).
- Palis, J. & Kingsley, P. D. Differential gene expression during early murine yolk sac development. *Mol. Reprod. Dev.* **42**, 19–27 (1995).
- Brownlie, A. *et al.* Positional cloning of the zebrafish *sauternes* gene: a model for congenital sideroblastic anaemia. *Nature Genet.* **20**, 244–250 (1998).

Acknowledgements

We thank J. Amatruda, C. Trenor, V. Sellers and J. Levy for critical review of this manuscript; P. Haffter and C. Nusslein-Volhard for providing the zebrafish blood mutants before publication; C. Amemiya, J. Postlethwait, D. Nathan, A. Oates and J. Best for helpful discussions and experimental advice; D. Giarla for administrative assistance; B. Hogan, J. Rossant and L. Solnica-Krezel for discussions on placental and yolk sac biology; and L. Kunkel, G. Gilliland and W. Talbot for support and advice. L.I.Z. and N.C.A. are Associate Investigators of the Howard Hughes Medical Institute. This work was supported by grants from the NIH.

Correspondence and requests for materials should be addressed to L.I.Z. (e-mail: zon@rascal.med.harvard.edu). cDNA sequences have been deposited in GenBank under accession codes AA226612 (zebrafish *ferroportin1*), AA226614 (human *ferroportin1*) and AA226613 (mouse *ferroportin1*).

Interaction between Wnt and TGF- β signalling pathways during formation of Spemann's organizer

Michiru Nishita^{*†}, Minako K. Hashimoto^{†‡}, Souichi Ogata[‡], Micheline N. Laurent^{‡§}, Naoto Ueno^{*}, Hiroshi Shibuya^{*} & Ken W. Y. Cho[‡]

^{*} Division of Morphogenesis, Department of Developmental Biology, National Institute for Basic Biology, Okazaki 444-8585, Japan

[‡] Department of Developmental and Cell Biology, University of California, Irvine, California 92617-2300, USA

[†] These authors contributed equally to this work

[§] Present address: Department of Biology, California Institute of Technology, 1200E. California Boulevard, Pasadena, California 91125, USA

Members of the Wnt and TGF- β superfamilies regulate both cell fate and proliferation during development and tissue maintenance^{1–3}. In the early amphibian embryo, the Wnt and TGF- β superfamily signalling cascades are required for the establishment of a dorsal signalling centre, Spemann's organizer^{4–8}. Intracellular proteins of both pathways, upon activation, translocate to the nucleus to participate in transcription. Here we show that β -catenin and Lef1/Tcf, which are downstream components of the Wnt signalling cascade, form a complex with Smad4, an essential mediator of signals initiated by members of the TGF- β growth factor superfamily. In *Xenopus*, this interaction directly and synergistically affects expression of the *twin* (*Xtwn*) gene during formation of the organizer. This is, to our knowledge, the first

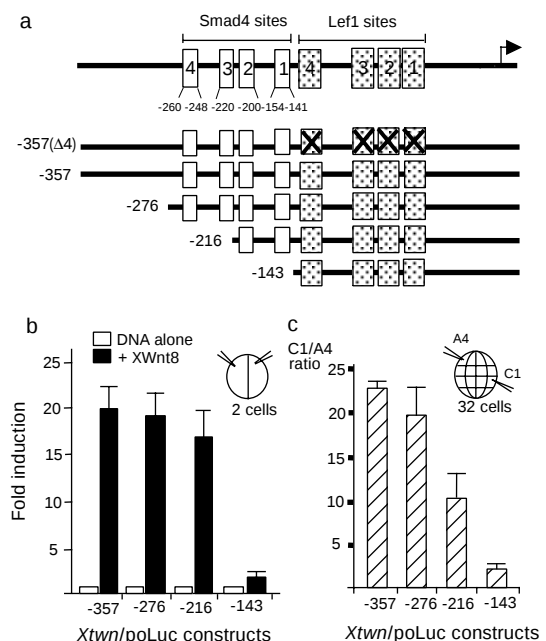


Figure 1 Deletion analysis of *Xtwn* reporter genes. **a**, Deletion constructs used for animal cap and blastomere injection assays. Putative Smad4 (open boxes: 5'-GTCT-3' (ref. 13)) and Lef1 (shaded boxes) binding sites are indicated. Sites 1, 2 and 3 are consensus Lef1/Tcf binding sites previously identified⁹, and site 4 is a newly identified non-consensus Lef1 site (AAGATCAAGGG; see Fig. 2d). Mutation of all four Lef1 binding sites generated -357(Δ4)*Xtwn*/Luc. Smad boxes 1, 2, 3 and 4 fall within the indicated three GST–Smad4MH1-protected regions. **b**, Deletion from -216 to -143 markedly reduces the Wnt responsiveness of the *Xtwn* promoter. **c**, Blastomere injection assay. A single A4 or C1 blastomere was injected at the 32-cell stage with the indicated deletion constructs. Fold induction relative to background (A4 levels of activity) is indicated.

demonstration of a physical interaction between TGF- β and Wnt signalling components *in vivo*.

The homeobox gene *Xtwn* and the closely related gene *siamois* (*Xsia*), both of which are expressed in the Spemann's organizer region, are direct targets of Wnt signalling^{4,8}. Their promoters contain essential consensus Lef1/Tcf binding sites, which mediate their responsiveness to Wnt. To identify other sequences that mediate Wnt responsiveness, we subcloned a series of 5' deletion constructs (Fig. 1a) into a luciferase reporter and injected them into four-cell embryos, either alone or with synthetic *Xwnt8* messenger RNA. At blastula stage the animal pole region was dissected away from the embryo and allowed to develop in culture. At early gastrula-equivalent stage, isolated explants were assayed for luciferase activity. *Xwnt8* induced all constructs at levels 15–20 times higher than reporter gene alone with the exception of –143*Xtwn*/Luc, which was induced 3–4-fold (Fig. 1b). The drop in Wnt responsiveness observed between –216 and –143*Xtwn*/Luc indicates that the Lef1/Tcf binding sites alone may be insufficient for a strong response to *Xwnt8* and that other elements in this region mediate Wnt signalling. Next we examined the responses of these constructs to endogenous Wnt signals using a blastomere injection assay. In *Xenopus*, a Wnt-like signal that stimulates *Xtwn* and *Xsia* is present in the dorsal region (future Spemann's organizer region) of the embryo^{4–8}. The C1 (prospective organizer) and A4 (prospective epidermis) blastomeres of 32-cell embryos were injected with the *Xtwn*/Luc deletion constructs. Embryos were collected at the early gastrula stage and luciferase activity was measured (Fig. 1c). Responsiveness to the endogenous Wnt-like signal decreased between –276 and –216*Xtwn*/Luc, and again between –216 and –143*Xtwn*/Luc, indicating that these regions may regulate expression of *Xtwn* *in vivo*. As the latter element is required for efficient Wnt stimulation in both animal cap assays (Fig. 1b) and blastomere injection assays (Fig. 1c), we conclude that the region between –216 and –143 is required, in addition to the Lef1/Tcf binding sites, for robust Wnt responsiveness.

The Smad2 pathway mediates activin/Vg1-like (TGF- β superfamily members) signalling in the early *Xenopus* embryo, and cooperates with Wnt signalling to induce expression of *Xsia* in *Xenopus*⁹. As these experiments used overexpression assays, we tested whether Smad4 is required for the *in vivo* induction of *Xtwn*. We microinjected dominant-negative Smad4 (DN-Smad4) mRNA into embryos to examine its effect on the induction of *Xtwn*. As Smad4 is required for all TGF- β superfamily signalling^{2,3,10,11}, overexpression of DN-Smad4 would be expected effectively to block all TGF- β -related signalling. Overexpression of DN-Smad4 blocked Wnt-mediated induction of the –357*Xtwn*/Luc reporter construct in animal caps (Fig. 2a). More importantly, overexpression of DN-Smad4 in the organizer inhibited the induction of –357*Xtwn*/Luc and endogenous *Xtwn* by the Wnt-like signal (Fig. 2b, and data not shown), implying that Smad4 is required for *Xtwn* induction. Two Smad2 mutants (P445H and D450E)¹² also reduced the transcriptional response of –357*Xtwn*/Luc (Fig. 2c).

We carried out DNaseI footprint analyses using glutathione S-transferase (GST)-purified full-length and MH1 domains of Smad1, 2, 3 and 4. Four Smad boxes fall within three DNaseI-protected regions (site 1, –154 to –141; sites 2 and 3, –220 to –200; site 4, –260 to –248) of the *Xtwn* promoter using the Smad4 MH1 domain (Fig. 2d, lanes 1–4 and 7). Two of these Smad boxes lie between –216 and –143, the region shown by deletion analysis to mediate Wnt responsiveness of the *Xtwn* promoter (Fig. 1b, c). One Smad4 binding site is immediately adjacent to a Lef1/Tcf binding site (compare lanes 7 to 9). Smad1, 2 and 3 (both full-length and MH1 domain proteins) did not bind strongly to the –357-base pair (bp) *Xtwn* promoter region (data not shown).

The proximity of a Smad4 binding site to a Lef1/Tcf binding site indicates that these signalling molecules may interact to regulate *Xtwn* expression. To address this issue, 293 cells were co-transfected

with Myc-tagged Smads and β -catenin, followed by immunoprecipitation of β -catenin using anti- β -catenin antibodies. The co-precipitating Smads were then visualized by western blot analyses using anti-Myc antibodies. β -Catenin specifically interacts with Smad4 (Fig. 3a). A similar experiment shows that Lef1 interacts with both haemagglutinin (HA)-tagged Smad4 and β -catenin (Fig. 3b).

If Smad4 interacts with Lef1/Tcf and β -catenin *in vivo*, Wnt signalling may trigger translocation of cytoplasmic Smad4 to the

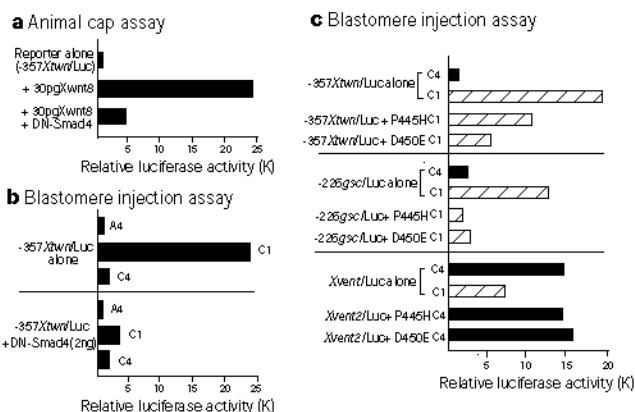


Figure 2 Requirement of Smad4 for *Xtwn* induction. **a**, Wnt induction of –357*Xtwn*/Luc is inhibited by DN-Smad4. –357*Xtwn*/Luc was injected either alone or with indicated mRNAs into two-cell embryos. Animal pole explants were collected at stage 8 and assayed for luciferase activity at stage 10.25. **b**, DN-Smad4 suppresses the –357*Xtwn*/Luc response to endogenous Wnt-like signals. **c**, DN-Smad2 mutant RNAs (1ng of P445H and D450E) interfere with –357*Xtwn*/Luc and –226*gsc*/Luc, but not *Xvent2*/Luc induction. **d**, DNaseI footprint analysis of the *Xtwn* promoter. MH1-Smad4-protected sites cover Smad box consensus sites of the *Xtwn* promoter as indicated (see also Fig. 1). Lef1-protected sites cover Lef1 consensus sites 2 and 3 and a non-consensus Lef1 binding site 4. NP, no protein.

phenotype²⁶. It will be interesting to investigate whether *Medea* is required for zygotic Wnt signalling. Cooperation of the TGF- β and Wnt signalling pathways may also mediate carcinogenesis. Progression of benign adenomatous polyps to invasive forms of carcinoma increases when members of both the TGF- β and Wnt signalling cascades are altered²⁷. The interaction of Lef1/Tcf with Smad4 adds additional layers of complexity to our expanding knowledge of the networks of regulation of the Wnt and TGF- β signalling cascades. However, convergence of these two signalling pathways may represent a simple and logical way to provide a greater diversity of cellular responses during development and tissue stasis.

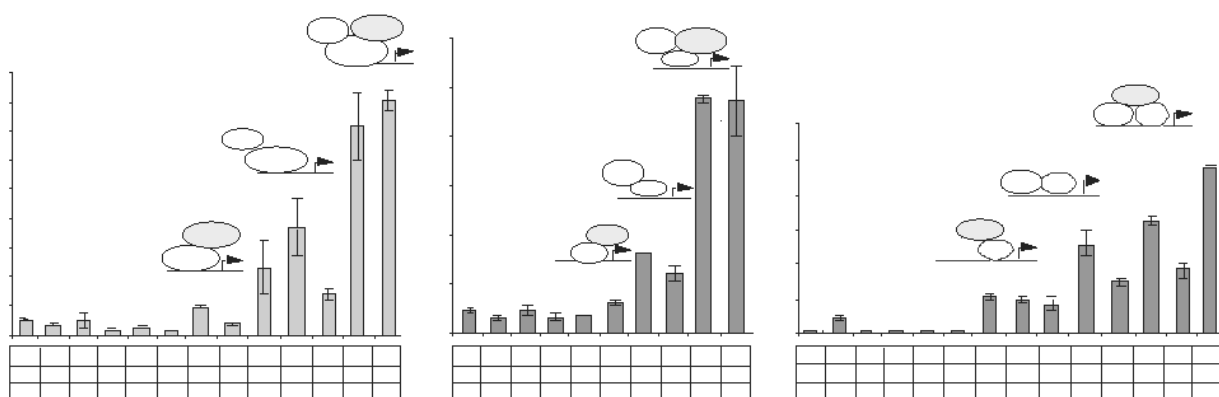
Methods

Embryo handling

Embryo manipulations and microinjections were as described^{7,28}.

Antibodies and immunoprecipitation

Antibodies to HA Y-11 (Santa Cruz), Myc 9E10 (Santa Cruz), Smad4 (Santa Cruz), Flag M2 (Kodak) and β -catenin (Transduction Laboratories and Sigma) were used to detect proteins. Immunoprecipitates and aliquots of total lysates were resolved by SDS-PAGE and transferred to PVDF membranes (Hybond-P, Amersham). Primary antibodies were visualized with HRP-conjugated antibodies to mouse or rabbit IgG using the Enhanced Chemiluminescence (ECL) Western Blotting System (Amersham).



Immunofluorescence

Whole-mount immunofluorescence in *Xenopus* embryos was carried out as described, with minor modifications²⁹. Briefly, embryos were injected anally at the 2-cell stage with *Xwnt8* RNA and fixed in 3.8% formaldehyde at stage 6. Animal caps were cut and stained with anti- β -catenin rabbit polyclonal (Sigma) and anti-Smad4 goat polyclonal antibodies (Santa Cruz), followed by FITC-conjugated donkey anti-rabbit IgG (Amersham) and HRP-conjugated donkey anti-goat IgG (Santa Cruz). To visualize the HRP-conjugated antibodies bound to anti-Smad4 antibodies, biotinylated tyramide and streptavidin-Texas Red were used according to the manufacturer's instructions (Tyramid Signal Amplification System, NEN). Image analysis was performed using a confocal laser-scanning microscope (Zeiss).

Transcriptional activity assay

Cells ((2–4) $\times 10^5$ cells per well) were seeded into 12-well plates. Indicated plasmids were transfected by the calcium phosphate precipitate method 24 h after seeding, and cells were collected and assayed for luciferase activity after 48 h. The total amount of DNA per transfection was equalized using empty vector DNA (pCS2+). A CMV5 β -gal expression construct was co-transfected to normalize for transfection efficiency.

Received 13 October; accepted 21 December 1999.

1. Cadigan, K. M. & Nusse, R. Wnt signaling: a common theme in animal development. *Genes Dev.* **11**, 3286–3305 (1997).
2. Heldin, C. H., Miyazono, K. & ten Dijke, P. TGF- β signalling from cell membrane to nucleus through SMAD proteins. *Nature* **390**, 465–471 (1997).
3. Massagué, J. TGF- β signal transduction. *Annu. Rev. Biochem.* **67**, 753–791 (1998).
4. Harland, R. & Gerhart, J. Formation and function of Spemann's organizer. *Annu. Rev. Cell Dev. Biol.* **13**, 611–667 (1997).
5. Brannon, M., Gomperts, M., Sumoy, L., Moon, R. T. & Kimelman, D. A β -catenin/XTcf-3 complex binds to the siamois promoter to regulate dorsal axis specification in *Xenopus*. *Genes Dev.* **11**, 2359–2370 (1997).
6. Moon, R. T. & Kimelman, D. From cortical rotation to organizer gene expression: toward a molecular explanation of axis specification in *Xenopus*. *BioEssays* **20**, 536–545 (1998).
7. Watabe, T. *et al.* Molecular mechanisms of Spemann's organizer formation: conserved growth factor synergy between *Xenopus* and mouse. *Genes Dev.* **9**, 3038–3050 (1995).
8. Laurent, M. N., Blitz, I. L., Hashimoto, C., Rothbacher, U. & Cho, K. W. The *Xenopus* homeobox gene *twint* mediates Wnt induction of *gooseoid* in establishment of Spemann's organizer. *Development* **124**, 4905–4916 (1997).
9. Crease, D. J., Dyson, S. & Gurdon, J. B. Cooperation between the activin and Wnt pathways in the spatial control of organizer gene expression. *Proc. Natl Acad. Sci. USA* **95**, 4398–4403 (1998).
10. Candia, A. F. *et al.* Cellular interpretation of multiple TGF- β signals: intracellular antagonism between activin/BVg1 and BMP-2/4 signaling mediated by Smads. *Development* **124**, 4467–4480 (1997).
11. Lagna, G., Hata, A., Hemmati-Brivanlou, A. & Massagué, J. Partnership between DPC4 and SMAD proteins in TGF- β signalling pathways. *Nature* **383**, 832–836 (1996).
12. Hoodless, P. A. *et al.* Dominant-negative Smad2 mutants inhibit activin/Vg1 signaling and disrupt axis formation in *Xenopus*. *Dev. Biol.* **207**, 364–379 (1999).
13. Shi, Y. *et al.* Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF- β signaling. *Cell* **94**, 585–594 (1998).
14. Yost, C. *et al.* The axis-inducing activity, stability, and subcellular distribution of β -catenin is regulated in *Xenopus* embryos by glycogen synthase kinase 3. *Genes Dev.* **10**, 1443–1454 (1996).
15. Waterman, M. L. & Jones, K. A. Purification of TCF-1 α , a T-cell-specific transcription factor that activates the T-cell receptor α gene enhancer in a context-dependent manner. *New Biol.* **2**, 621–636 (1990).
16. Fujita, T., Nolan, G. P., Liou, H. C., Scott, M. L. & Baltimore, D. The candidate proto-oncogene *bcl-3* encodes a transcriptional coactivator that activates through NF- κ B p50 homodimers. *Genes Dev.* **7**, 1354–1363 (1993).
17. McKendry, R., Hsu, S. C., Harland, R. M. & Grosschedl, R. LEF-1/TCF proteins mediate wnt-inducible transcription from the *Xenopus* nodal-related 3 promoter. *Dev. Biol.* **192**, 420–431 (1997).
18. He, T. C. *et al.* Identification of c-MYC as a target of the APC pathway. *Science* **281**, 1509–1512 (1998).
19. Tetsu, O. & McCormick, F. β -catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* **398**, 422–426 (1999).
20. Christian, J. L. & Moon, R. T. Interactions between *Xwnt-8* and Spemann organizer signaling pathways generate dorsoventral pattern in the embryonic mesoderm of *Xenopus*. *Genes Dev.* **7**, 13–28 (1993).
21. Jones, C. M., Lyons, K. M., Lapan, P. M., Wright, C. V. & Hogan, B. L. DVR-4 (bone morphogenetic protein-4) as a posterior-ventralizing factor in *Xenopus* mesoderm induction. *Development* **115**, 639–647 (1992).
22. Hoppner, S. & Moon, R. T. BMP-2/-4 and Wnt-8 cooperatively pattern the *Xenopus* mesoderm. *Mech. Dev.* **71**, 119–129 (1998).
23. Riese, J. *et al.* LEF-1, a nuclear factor coordinating signaling inputs from wingless and decapentaplegic. *Cell* **88**, 777–787 (1997).
24. Theisen, H., Haerry, T. E., O'Connor, M. B. & Marsh, J. L. Developmental territories created by mutual antagonism between Wingless and Decapentaplegic. *Development* **122**, 3939–3948 (1996).
25. Brook, W. J. & Cohen, S. M. Antagonistic interactions between wingless and decapentaplegic responsible for dorsal-ventral pattern in the *Drosophila* leg. *Science* **273**, 1373–1377 (1996).
26. Wisotzkey, R. G. *et al.* Medea is a *Drosophila* Smad4 homolog that is differentially required to potentiate DPP responses. *Development* **125**, 1433–1445 (1998).
27. Takaku, K. *et al.* Intestinal tumorigenesis in compound mutant mice of both Dpc4 (Smad4) and Apc genes. *Cell* **92**, 645–656 (1998).
28. Cho, K. W., Blumberg, B., Steinbeisser, H. & De Robertis, E. M. Molecular nature of Spemann's organizer: the role of the *Xenopus* homeobox gene *gooseoid*. *Cell* **67**, 1111–1120 (1991).
29. Larabell, C. A. *et al.* Establishment of the dorso-ventral axis in *Xenopus* embryos is presaged by early asymmetries in β -catenin that are modulated by the Wnt signaling pathway. *J. Cell. Biol.* **136**, 1123–1136 (1997).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank I. Blitz, M. Waterman and K. Matsumoto for their critical reading of the manuscript and helpful comments; S. Nagai, Y. Takahashi and C. Nugas for assistance; R. Moon, P. ten Dijke, R. Derynck, L. Attisano, O. Tetsu, F. McCormick, J. Thomsen, T. Akiyama, M. Prieve and T. Li for reagents; and L. Rafferty for discussion. This work was supported by an NIH grant and by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan and 'Research for the Future' program of the Japan Society for the promotion of Science.

Correspondence and requests for materials should be addressed to H.S. (e-mail: shibuya@nibb.ac.jp) or K.W.Y.C. (e-mail: kwcho@uci.edu).

Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis

Sha Mi, Xinhua Lee, Xiang-ping Li, Geertruida M. Veldman, Heather Finnerty, Lisa Racie, Edward LaVallie, Xiang-Yang Tang, Philippe Edouard, Steve Howes, James C. Keith Jr & John M. McCoy*

Genetics Institute, Inc., 87 CambridgePark Drive, Cambridge, Massachusetts 02140, USA

* Present address: Biogen Inc., 14 Cambridge Center, Cambridge, Massachusetts 02142, USA

Many mammalian viruses have acquired genes from their hosts during their evolution¹. The rationale for these acquisitions is usually quite clear: the captured genes are subverted to provide a selective advantage to the virus. Here we describe the opposite situation, where a viral gene has been sequestered to serve an important function in the physiology of a mammalian host. This gene, encoding a protein that we have called syncytin, is the envelope gene of a recently identified human endogenous defective retrovirus, HERV-W². We find that the major sites of syncytin expression are placental syncytiotrophoblasts, multinucleated

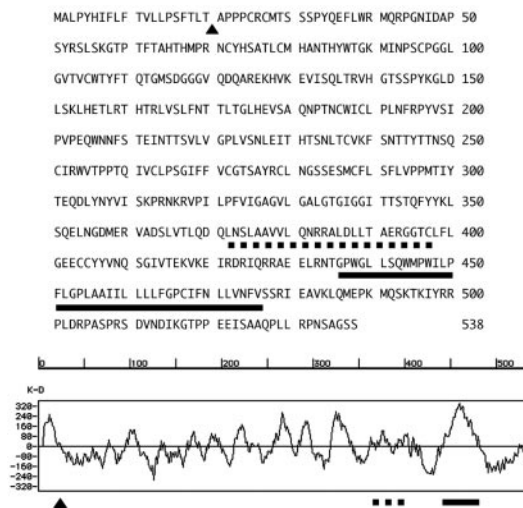


Figure 1 Primary sequence and hydrophobicity plot of human syncytin. Arrowhead, predicted signal sequence cleavage site. Thick broken line, putative immunosuppressive region. On the basis of hydrophobicity analysis (lower panel), syncytin is predicted to be a type I membrane protein; its transmembrane region is delineated by a thick solid line.

cells that originate from fetal trophoblasts. We show that expression of recombinant *syncytin* in a wide variety of cell types induces the formation of giant syncytia, and that fusion of a human trophoblastic cell line expressing endogenous *syncytin* can be inhibited by an anti-*syncytin* antiserum. Our data indicate that *syncytin* may mediate placental cytotrophoblast fusion *in vivo*, and thus may be important in human placental morphogenesis.

A complementary DNA fragment comprising the 5' end of the *syncytin* gene was isolated from a human testis library as part of a project to identify novel secreted proteins using the yeast signal sequence trap³. A full-length human *syncytin* cDNA was subsequently isolated from the same source: its sequence and features are shown in Fig. 1.

Protein homology searches of public databases using the *syncytin* protein as the query reveal significant similarity to many retroviral envelope proteins, in particular to the *env* proteins encoded by a newly identified human endogenous retrovirus, HERV-W² (GenBank accession no. AAD14546, 100% identity), a baboon endogenous retrovirus (GenBank accession no. D10032, 34% identity) and

SRV-2 (GenBank accession no. M16605, 33% identity). A search of DNA databases using *syncytin* cDNA as the query returns identities to multiple expressed sequence tags, providing additional evidence that the human *syncytin* gene is expressed. DNA searches also reveal the existence of several sequences closely related to *syncytin* on human chromosomes 4, 6, 7, 16, 17, 21, 22 and X. The chromosome 7 sequence (GenBank accession no. AC000064, map position 7q21–7q22), containing an intact *syncytin* ORF and possessing 100% identity to the *syncytin* cDNA, is the best candidate for the true *syncytin* gene. Flanking sequences reveal the presence of a defective provirus that has been named as the prototype of a new endogenous retrovirus family, HERV-W (ref. 2). *Syncytin* is the envelope protein of this endogenous retrovirus.

DNA from nine species was examined by Southern blot, probing with *syncytin*-coding cDNA under stringent conditions. We identified multiple hybridizing bands in human DNA (Fig. 2b), confirming our public sequence database findings. Multiple copies of *syncytin*-related sequences were also observed in rhesus monkey DNA; however, no evidence was found, under the experimental conditions used, for homologues in any other organisms tested, including representatives of several other mammalian orders. We estimate that there are at least 15 copies of sequences closely related to *syncytin* in the human genome.

Syncytin gene expression was examined in 23 human tissues by northern blot (Fig. 2a). Robust expression was observed in placenta, with two major transcripts of 4 and 8 kilobases (kb). Weaker expression was seen in testis, with the 8-kb transcript predominating. We did not detect *syncytin* expression in any other human tissue. Multiple *syncytin* cDNA clones subsequently isolated from a human placental library were identical to the original testis isolate in the *syncytin*-coding region.

In situ hybridizations (Fig. 3a,b) performed on tissue sections prepared from the villous region of human placenta revealed that *syncytin* expression is restricted to syncytiotrophoblasts, a fused cell type of fetal origin. Syncytiotrophoblasts arise by cell fusion from

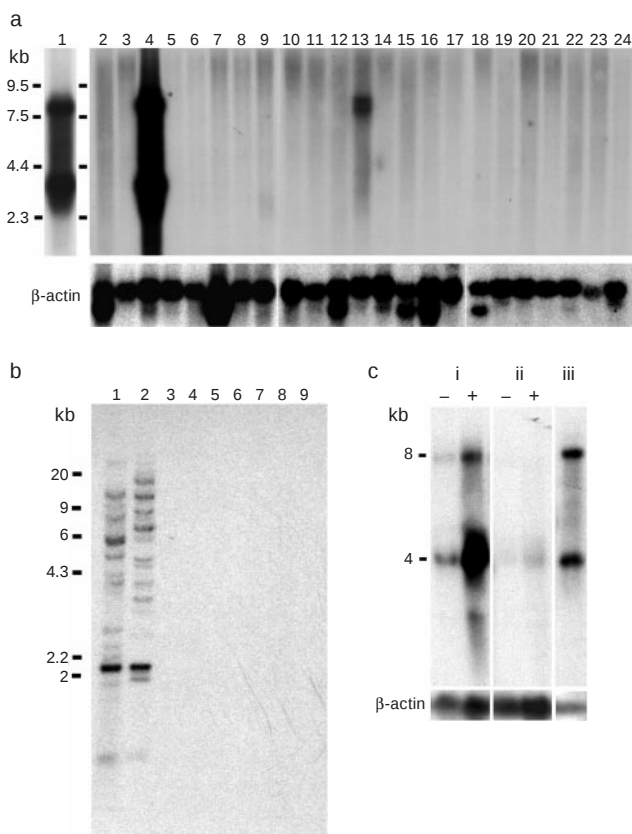


Figure 2 *Syncytin* gene distribution and expression. **a**, Northern blots showing *syncytin* expression in human tissues. Poly(A)⁺ RNA from 23 tissues were probed with the *syncytin* coding region. Lane 1: placenta (short exposure); 2: heart; 3: brain; 4: placenta; 5: lung; 6: liver; 7: skeletal muscle; 8: kidney; 9: pancreas; 10: spleen; 11: thymus; 12: prostate; 13: testis; 14: ovary; 15: small intestine; 16: colon; 17: peripheral blood leukocyte; 18: stomach; 19: thyroid; 20: spinal cord; 21: lymph node; 22: trachea; 23: adrenal; 24: bone marrow. The β -actin loading control is shown below. **b**, Southern blot of genomic DNA from nine species probed with *syncytin* coding-region cDNA. Lane 1: human; 2: rhesus monkey; 3: rat; 4: mouse; 5: dog; 6: cow; 7: rabbit; 8: chicken; 9: yeast. **c**, Northern blot of *syncytin* expression in choriocarcinoma cell lines probed with the *syncytin* coding region. RNA was derived from (i) BeWo choriocarcinoma cells ($\pm$ forskolin treatment); (ii) JEG3 choriocarcinoma cells ($\pm$ forskolin treatment); and (iii) human term placenta. The β -actin loading control is shown; estimates of fold inductions were corrected for loading.

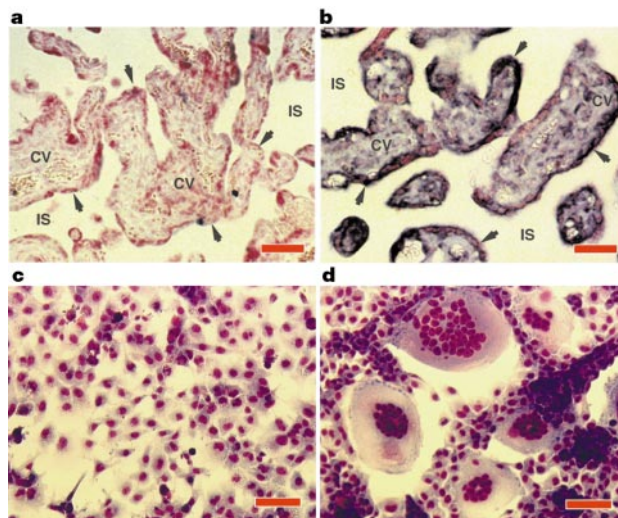


Figure 3 *In situ* hybridizations and *syncytin*-mediated COS cell fusion. **a, b**, *In situ* hybridizations of sections taken through the villous region of human term placenta. **a**, Negative control, section probed with digoxigenin-labelled sense-strand *syncytin* RNA. **b**, Section probed with digoxigenin-labelled antisense-strand *syncytin* RNA. Arrowheads, syncytiotrophoblast layer; CV, chorionic villi (fetal tissue); IS, intervillous space (maternal blood space). **c, d**, *Syncytin*-mediated cell fusion. **c**, COS cells transfected with a vector containing the *syncytin* gene in the reverse orientation for expression (negative control). **d**, COS cells transfected with a vector containing the *syncytin* gene in the sense orientation. Red scale bars, 100 μ m.

cytotrophoblasts and constitute the boundary layer between maternal and fetal tissue. They are important in maternal–fetal exchange, in tissue remodelling during placental development and in protecting the developing fetus from the maternal immune response^{4–6}. As retroviral *env* proteins are often involved in promoting cell fusion events, we investigated whether syncytin could mediate fusion in defined *in vitro* culture systems. Transfection of COS cells with a *syncytin* expression plasmid resulted in the generation of large COS syncytia (Fig. 3c,d), often containing more than 30 aggregated nuclei and with each aggregate surrounded by an extensive area of thin, extended cytoplasm. To prove that these COS syncytia arose through fusion and not by defective cell division, we devised a two-component fusion assay. Human interleukin 12 (IL-12) is a heterodimeric cytokine comprising disulphide-linked subunits with relative molecular masses (M_r) of 35K (p35) and 40K (p40). Both subunits are required for bioactivity. We performed mixing experiments using COS cells co-transfected in different combinations with expression plasmids carrying the genes for p35, p40 and *syncytin* (in both forward and reverse orientations). IL-12 activity in the resulting conditioned medium was used as an indicator of cell fusion. Table 1 summarizes the results. IL-12 activity was detected only if p35 and p40 were co-transfected into the same batch of COS cells (the positive control), or if cells co-transfected with p40 and *syncytin* were mixed with cells transfected with p35. The latter observation confirmed that syncytin mediates true cell fusion. During these experiments, levels of IL-12 generated by fused cells were consistently higher than those seen in the p35/p45 co-transfection

controls. The reason for this remains unclear. IL-12 activity was seen regardless of whether the batch of COS cells transfected with p35 was also co-transfected with *syncytin*, indicating that homophilic interactions are not required for syncytin-mediated fusion in COS cells. Furthermore, we have found that a wide variety of cell types can fuse to *syncytin*-transfected COS cells, including human (HeLa) and rodent (CHO) lines. Cell fusion was also observed in an insect line (Sf9) expressing *syncytin* (data not shown). To rule out the possibility of a ubiquitously expressed syncytin-‘binding protein’ acting *in trans* with syncytin to mediate cell fusion, we investigated whether syncytin alone could mediate fusion of COS cells to simple lipid vesicles. Using human heart lipid extracts we generated liposomes enclosing a solution of a green fluorescent protein (GFP) expression plasmid. This liposome preparation was added separately to COS cells transfected with a *syncytin* expression vector, and to COS cells transfected with an expression vector carrying the *syncytin* gene in the reverse orientation (the negative control). We observed GFP fluorescence only in the COS cells fused as a consequence of *syncytin* expression (data not shown), indicating that syncytin acting alone can mediate membrane fusion. Although this ability is intriguing, it is not unprecedented⁷.

BeWo, a human trophoblastic choriocarcinoma line⁸, can be induced by forskolin to differentiate *in vitro* into fused cells resembling syncytiotrophoblasts⁹. Figure 2c shows that *syncytin* transcription in BeWo cells increases at least fivefold in response to treatment with forskolin, correlating well with the increased BeWo cell fusion seen under these conditions. In contrast, a second

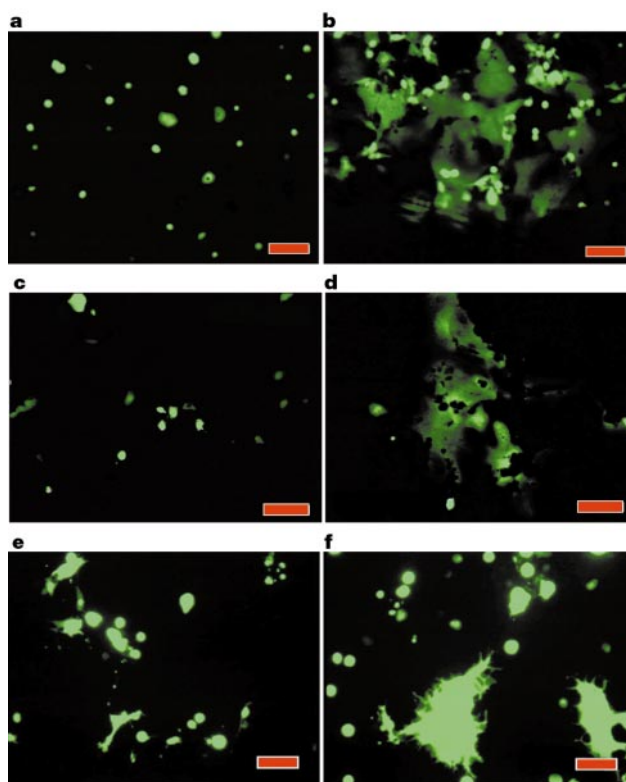


Figure 4 Fusion of BeWo choriocarcinoma cells to GFP-labelled COS cells. **a,b**, BeWo cells fuse in response to forskolin treatment. BeWo choriocarcinoma cells cultured in the absence (**a**) or presence (**b**) of forskolin were co-cultured for 48 h with COS cells transfected with a GFP expression plasmid. The small, circular, intensely fluorescent green cells are unfused COS cells. The large, irregular, less intensely fluorescent cells are BeWo cells fused to each other and to GFP-expressing COS cells. In the absence of forskolin <5% of the BeWo cells were fused; in the presence of forskolin 80–90% of BeWo cells were fused. **c,d**, *Syncytin* transfection is sufficient to mediate BeWo cell fusion. BeWo cells were transfected with either a *syncytin* expression vector (**d**) or with a

vector containing the *syncytin* gene in the reverse orientation for expression (**c**) and co-cultured for 48 h with GFP-expressing COS cells in the absence of forskolin. 50–60% of the BeWo cells were fused by *syncytin* transfection. **e,f**, An anti-*syncytin* antiserum can reduce forskolin-induced BeWo cell fusion. BeWo choriocarcinoma cells were co-cultured in the presence of forskolin for 48 h with COS cells transfected with a GFP expression plasmid. Co-culture was performed either in the presence of a 1:20 dilution of a rabbit anti-*syncytin* antiserum (**e**) or a 1:20 dilution of rabbit pre-immune serum (**f**). The antiserum inhibited the formation of fluorescent fused cells by 49% ($\pm$ 9%, six determinations). Red scale bars, 200 μ m.

choriocarcinoma line which does not fuse in response to forskolin, JEG3 (ref. 10), showed only low *syncytin* expression. Although this increased slightly following forskolin treatment, overall *syncytin* transcription remained lower than in untreated BeWo cells.

COS cells were transfected with a GFP expression vector and co-cultured with BeWo cells previously treated or untreated with forskolin (Fig. 4a,b). Many more large, fused, green fluorescent cells were observed in the culture treated with forskolin (Fig. 4b) than in the non-forskolin-treated control (Fig. 4a), indicating that the forskolin-induced BeWo cells had fused to the GFP-containing COS cells. A similar increase in BeWo/COS cell fusion could be achieved in the absence of forskolin if BeWo cells were first transfected with a *syncytin* expression vector (Fig. 4d), as compared to a vector-transfected control (Fig. 4c), indicating that *syncytin* expression is sufficient to produce BeWo/COS cell fusion. To show that increased BeWo/COS cell fusion in response to forskolin treatment was a direct result of increased endogenous *syncytin* expression in BeWo cells, the experiment described above (Fig. 4b) was repeated in the presence of either a 1:20 dilution of a rabbit anti-*syncytin* antiserum (Fig. 4e) or a 1:20 dilution of pre-immune rabbit serum (Fig. 4f). The anti-*syncytin* antiserum reduced cell fusion significantly (to $49 \pm 9\%$), indicating that *syncytin* is a major agent responsible for forskolin-induced BeWo/COS cell fusion, although at present we cannot rule out the potential involvement of additional proteins that may regulate *syncytin* activity *in vivo*.

There have been previous reports of endogenous retroviral gene expression in mammals^{11–13}, but only a few examples have clearly shown this expression to be important in host biology^{14,15}. Repeated observations of retroviral-like particles in placenta¹⁶ and the existence of fused placental cells morphologically reminiscent of virally induced syncytia led to the proposal that an ancient retroviral infection may have been a pivotal event in mammalian evolution¹⁷. Our work provides some support for this hypothesis; however, fused cell types are found in the placentas of many mammals¹⁸, and yet we detected *syncytin*-related sequences only in humans and in the one primate species we examined. The reason for this discrepancy is unclear.

The placenta is a secretory exchange organ, with syncytiotrophoblasts comprising the exchanging surface. Assuming that there are no changes in cytoplasmic volume or total membrane surface area as cells fuse, geometry dictates that a growing syncytium will increasingly possess excess cell membrane above that needed simply to enclose cell contents. Thus, syncytium formation in placenta may be a mechanism for providing an extended surface for maternal–fetal exchange. Moreover, our preliminary observations of increased IL-12 secretion in *syncytin*-fused COS cells may indicate that fused cells have enhanced secretory abilities. *Syncytin* may also be involved in directing placental morphogenesis. Trophoblast invasion of the maternal decidua and myometrium is a carefully orchestrated process, with insufficient infiltration of the

uterine wall implicated in placental disorders such as pre-eclampsia¹⁹, and uncontrolled trophoblast invasion observed in choriocarcinoma and other gestational trophoblastic diseases²⁰. We observe a roughly fivefold decrease in DNA synthesis following treatment of BeWo cells with forskolin, and see a similar decrease compared to controls for COS cells transfected with a *syncytin* expression plasmid (data not shown). *Syncytin* may thus also limit invasiveness by inducing trophoblast fusion and terminal differentiation.

Many retroviral *env* proteins are reported to mediate immunosuppression through a conserved segment of 25 amino acids located in their extracellular domains²¹. This sequence is also found in *syncytin* (amino-acid residues 373–397, Fig. 1), and thus abundant *syncytin* expression in placental syncytiotrophoblasts may contribute towards immune tolerance of the developing embryo, although additional mechanisms have also been proposed^{6,22,23}.

In humans there exists a second endogenous retrovirus, ERV-3 (ref. 24), for which a role in placental function has been suggested^{13,25}. ERV-3 bears many striking similarities to the *syncytin* provirus, HERV-W. Like HERV-W, ERV-3 maps to the long arm of chromosome 7, prominent ERV-3 transcripts of 3.5 and 9 kb are observed in placenta and weaker ERV-3 expression is seen in testis. Notwithstanding these similarities it is clear on closer examination that ERV-3 *env* and *syncytin* are very different. ERV-3 *env* has no obvious transmembrane region, and shares primary sequence homology with *syncytin* only in the immunosuppressive peptide region discussed above. An increase in ERV-3 *env* expression was correlated with differentiation and increased cell fusion in forskolin-treated or ERV-3-transfected BeWo cells²⁶. To investigate whether ERV-3 *env*, like *syncytin*, could directly mediate cell fusion in COS cells, we isolated and characterized several full-length ERV-3 *env* cDNAs from human placenta and transfected them into COS cells. Although equivalent levels of radiolabelled *syncytin* and ERV-3 *env* production could be observed on protein gels, ERV-3 *env* did not mediate discernible cell fusion in these experiments (data not shown), in contrast to *syncytin* transfections (Fig. 3d).

In conclusion, we have identified a captured retroviral gene, *syncytin*, which may be important in human placental biology. The possibility that *syncytin* dysregulation or dysfunction is involved in pathologies such as preeclampsia or choriocarcinoma is a focus of ongoing studies. We hope that a clearer understanding of *syncytin* gene regulation and structure/function may lead to new approaches in the diagnosis, prevention or treatment of these diseases.

Methods

Chemicals and reagents

Lipofectamine was purchased from Life Technologies. Forskolin was obtained from Sigma. Certain DNA and RNA blots were purchased from Clontech. Digoxigenin NTP mix was from Boehringer Mannheim. Human heart lipid extract was from Avanti Polar Lipids. 4-Nitro blue tetrazolium chloride (NBT) and X-phosphate/5-bromo-4-chloro-3-indolyl-phosphate (BCIP) were from Promega. Diff-Quick stain kit was from Dade International Inc. Rabbit anti-*syncytin* antiserum was raised against a mixture of six KLH-conjugated *syncytin* peptides (residues 29–44, 50–65, 69–74, 301–315, 344–358 and 373–397).

Cell lines and placental tissues

Human choriocarcinoma cell lines BeWo and JEG3 were obtained from ATCC. BeWo cells were grown in FK12, JEG3 cells in MEM and COS-1 cells in DMEM. All media contained 10% fetal bovine serum (Sigma), 2 mM L-glutamine and 1% penicillin G-streptomycin (Life Technologies). Treatments of choriocarcinoma lines with 50 μ M forskolin were performed at 50% confluence. Fresh human placental tissue was obtained from Advance Bioscience Resources Inc.

COS-1 cell transfections

COS-1 cells were grown to 60% confluence in P100 dishes. We performed transfections with expression plasmids²⁷ using 500 μ l serum-free medium containing 50 μ l lipofectamine and 8 μ g total DNA. After 30 min incubation of this mixture at room temperature, 3.5 ml serum-free medium was added before mixing with cells. Cultures were inspected for cell fusion 48 h after transfection.

Table 1 Functional interleukin-12 expression as an assay for *syncytin*-mediated cell fusion

COS-cell transfection 1	COS-cell transfection 2	IL-12 (ng ml ⁻¹)
p40.p35	–	2,000
<i>syncytin</i> /p40	<i>syncytin</i> /p35	8,000
<i>syncytin</i> /p40	<i>syncytin</i> -reverse/p35	9,500
<i>syncytin</i> -reverse/p40	<i>syncytin</i> -reverse/p35	<10
p40	p35	<10
<i>syncytin</i> /p40	–	<10
<i>syncytin</i> -reverse/p40	–	<10
<i>syncytin</i> /p35	–	<10
<i>syncytin</i> -reverse/p35	–	<10
p40	–	<10
p35	–	<10

Interleukin-12 levels in conditioned medium were estimated from the standard curve of a human PHA blast proliferation assay.

Blot hybridizations

For the Southern blot, 4 µg of genomic DNA from each of nine eukaryotes was digested with *EcoRI*, separated on a 0.7% agarose gel, transferred onto nylon membrane and hybridized overnight at 65 °C in 6 × SSC/0.5% SDS using a ³²P-radiolabelled *syncytin* coding region probe (200 base pairs). The filter was washed in 2 × SSC/0.5% SDS and in 0.2 × SSC/0.1% SDS, both for 2 h at 65 °C, before X-ray film exposure. For northern blots hybridizations and washes were as described above, but at 72 °C.

In situ hybridization of placental tissue

Sections were prepared and processed as described²⁸, and were probed with digoxigenin-labelled *syncytin* antisense and sense RNA²⁹. Sections were stained using the alkaline phosphate substrates NBT and BCIP using conditions recommended by the manufacturer.

Fused COS-1 cell staining

COS-1 cells were stained 48 h after transfection. Cells were washed with PBS, air dried, fixed with Diff-Quick fixative (10 min), treated with Diff-Quick solution I (10 min), washed again with PBS, stained with Diff-Quick solution II (2 min), washed again with PBS and visualized under a light microscope.

IL-12 PHA blast assay

PHA (phytohaemagglutinin) blasts were prepared by stimulating human peripheral blood cells with PHA for four days, followed by three days with PHA and IL-12, and were then used to measure bioactive IL-12 in a proliferation assay³⁰.

Liposome cell fusion

We mixed 2 mg heart lipid extract with 5 µg GFP expression plasmid in 1 ml PBS. The mixture was homogenized (10 min) with a hand-held homogenizer and centrifuged (30 min) at 8,000g. The pellet, resuspended in PBS, was added to COS cells expressing *syncytin*. GFP expression was detected after 48 h by fluorescence microscopy.

Received 29 April; accepted 20 December 1999.

- Ploegh, H. L. Viral strategies of immune evasion. *Science* **280**, 248–253 (1998).
- Blond, J.-L. *et al.* Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. *J. Virol.* **73**, 1175–1185 (1999).
- Jacobs, K. A. *et al.* A genetic selection for isolating cDNAs encoding secreted proteins. *Gene* **198**, 289–296 (1997).
- Cross, J. C., Werb, Z. & Fisher, S. J. Implantation and the placenta: key pieces of the development puzzle. *Science* **266**, 1508–1518 (1996).
- Lala, P. K. & Hamilton, G. S. Growth factors, proteases and protease inhibitors in the maternal fetal dialog. *Placenta* **17**, 545–555 (1996).
- Munn, D. H. *et al.* Prevention of allogeneic fetal rejection by tryptophan catabolism. *Science* **281**, 1191–1193 (1998).
- Yamada, S. & Ohnishi, S. Vesicular stomatitis virus binds and fuses with phospholipid domain in target cell membranes. *Biochemistry* **25**, 3703–3708 (1986).
- Pattillo, R. A., Gey, G. O., Delfs, E. & Mattingly, R. F. Human hormone production *in vitro*. *Science* **159**, 1467–1469 (1968).
- Lyden, T. W., Ng, A.-K. & Rote, N. S. Modulation of phosphatidylserine epitope expression by BeWo cells during forskolin treatment. *Placenta* **14**, 177–186 (1993).
- Kohler, P. O. & Bridson, W. E. Isolation of hormone-producing clonal lines of human choriocarcinoma. *J. Clin. Endocrinol. Metab.* **32**, 683–687 (1971).
- Larsson, E. *et al.* Expression of an endogenous retrovirus, HERV-R, in human tissues. *J. Cancer Res. Clin. Oncol.* **119** (Suppl. 1), 6 (1993).
- Boyd, M. T., Bax, C. M., Bax, B. E., Bloxam, D. L. & Weiss, R. A. The human endogenous retrovirus ERV-3 is upregulated in differentiating placental trophoblast cells. *Virology* **196**, 905–909 (1993).
- Venables, P. J., Brookes, S. M., Griffiths, D., Weiss, R. A. & Boyd, M. T. Abundance of an endogenous retroviral envelope protein in placental trophoblasts suggests a biological function. *Virology* **211**, 589–592 (1995).
- Limjoco, T. I., Dickie, P., Ikeda, H. & Silver, J. Transgenic Fv-4 mice resistant to Friend virus. *J. Virol.* **67**, 4163–4168 (1993).
- Best, S., Le Tissier, P., Towers, G. & Stoye, J. P. Positional cloning of the mouse retrovirus restriction gene Fv1. *Nature* **382**, 826–829 (1996).
- Kalter, S. S. *et al.* A comparative study on the presence of C-type viral particles in placentas from primates and other animals. *Bibl. Haematol.* **40**, 391–401 (1975).
- Harris, J. R. The evolution of placental mammals. *FEBS Lett.* **295**, 3–4 (1991).
- Pijnenborg, R., Robertson, W. B., Brosens, I. & Dixon, G. Trophoblast invasion and the establishment of haemochorial placentation in man and laboratory animals. *Placenta* **2**, 71–92 (1981).
- Fox, H. in *Pathology of the Placenta* Chapter 6 (ed. Fox, H.) 151–175 (W. B. Saunders, Philadelphia, 1997).
- Baergen, R. N. Gestational choriocarcinoma. *Gen. Diagn. Pathol.* **143**, 127–141 (1997).
- Cianciolo, G. J., Copeland, T. D., Oroszlan, S. & Snyderman, R. Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. *Science* **230**, 453–455 (1985).
- Kovats, S. *et al.* A class I antigen, HLA-G, expressed in human trophoblasts. *Science* **248**, 220–223 (1990).
- Hunt, J. S. Immunobiology of pregnancy. *Curr. Opin. Immunol.* **4**, 591–596 (1992).
- Cohen, M., Powers, M., O'Connell, C. & Kato, N. The nucleotide sequence of the env gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. *Virology* **147**, 449–458 (1985).
- Larsson, E., Andersson, A. C. & Nilsson, B. O. Expression of an endogenous retrovirus (ERV3 HERV-R) in human reproductive and embryonic tissues—evidence for a function for envelope gene products. *Ups. J. Med. Sci.* **99**, 113–120 (1994).

- Lin, L., Xu, B. & Rote, N. S. Expression of endogenous retrovirus ERV-3 induces differentiation in BeWo, a choriocarcinoma model of human placental trophoblast. *Placenta* **20**, 109–118 (1999).
- Kaufman, R. J., Davies, M. V., Wasley, L. C. & Michnick, D. Improved vectors for stable expression of foreign genes in mammalian cells by use of the untranslated leader sequence from EMC virus. *Nucleic Acids Res.* **19**, 4485–4490 (1991).
- Sassoon, D. & Rosenthal, N. Detection of messenger RNA by *in situ* hybridization. *Methods Enzymol.* **225**, 384–404 (1993).
- Fischer, D., Weisenberger, D. & Scheer, U. Assigning functions to nucleolar structures. *Chromosoma* **101**, 133–140 (1991).
- Gately, M. K., Chizzonite, R. & Presky, D. H. in *Current Protocols in Immunology* **6**, 16 (eds Coligan, J. E., Kruisbeek, A. M., Margulies, D. H., Shevach, E. M. & Strober, W.) (Wiley, New York, 1996).

Acknowledgements

We thank the Genetics Institute signal sequence trap team for the initial cloning of *syncytin*; J. Wooters for help in liposome preparations; B. Gimlich, I. Moutsatsos and the Genetics Institute Developmental Biology group for microscopy assistance; the Genetics Institute DNA synthesis group for oligonucleotides; and M. Davies, R. Pijnenborg and K. Turner for critical review of this manuscript.

Correspondence should be directed to J.M.M. (e-mail john_mccoy@biogen.com) or S.M. (e-mail smi@genetics.com). Requests for materials should be directed to S.M. The sequence of the *syncytin* cDNA used in this work can be found in GenBank, accession no. AF208161.

naked cuticle encodes an inducible antagonist of Wnt signalling

Wenlin Zeng^{*†‡}, Keith A. Wharton Jr^{*†§}, Judith A. Mack^{*†‡}, Kevin Wang^{†‡}, Matthew Gadbaw^{†‡}, Kaye Suyama^{†‡}, Peter S. Klein^{||} & Matthew P. Scott^{†‡}

Departments of [†]Developmental Biology, [‡]Genetics, and [§]Pathology, Howard Hughes Medical Institute, Beckman Center B300, 279 Campus Drive, Stanford University School of Medicine, Stanford, California 94305, USA ^{||}Department of Internal Medicine, Howard Hughes Medical Institute, University of Pennsylvania School of Medicine, Clinical Research Building, Room 405, 415 Curie Boulevard, Philadelphia, Pennsylvania 19104-6148, USA ^{*} These three authors contributed equally to this work

During animal development, cells have to respond appropriately to localized secreted signals. Proper responses to Hedgehog, transforming growth factor-β, epidermal growth factor and fibroblast growth factor/Ras signals require cognate inducible antagonists such as Patched, Dad, Argos and Sprouty¹. Wnt signals are crucial in development and neoplasia². Here we show that *naked cuticle* (*nkd*), a *Drosophila* segment-polarity gene, encodes an inducible antagonist for the Wnt signal Wingless (Wg). In fly embryos and imaginal discs *nkd* transcription is induced by Wg. In embryos, decreased *nkd* function has an effect similar to excess Wg; at later stages such a decrease appears to have no effect. Conversely, overproduction of Nkd in *Drosophila* and misexpression of Nkd in the vertebrate *Xenopus laevis* result in phenotypes resembling those of loss of Wg/Wnt function. *nkd* encodes a protein with a single EF hand (a calcium-binding motif) that is most similar to the recoverin family of myristoyl switch proteins. Nkd may therefore link ion fluxes to the regulation of the potency, duration or distribution of Wnt signals. Signal-inducible feedback antagonists such as *nkd* may limit the effects of Wnt proteins in development and disease.

Wg is critical for patterning events during the three stages of embryonic segmentation in *Drosophila*³. First, 3–3.5 h after egg laying (AEL), adjacent stripes of cells produce Wg and Hedgehog (Hh). Engrailed (En) and Hh are produced in the same cells. Second, between 3.5 and 6 h AEL, Wg maintains *hh/en* transcription and Hh maintains *wg* transcription, producing a transient segmentation landmark, the parasegmental groove. Finally, from 6 h AEL

on, definitive segmentation results in ventral epidermal cells synthesizing a segmentally repeated trapezoidal array of six unique types of cell protrusion called denticles, interspersed with naked (denticle-free) cuticle. During this stage, Wg specifies naked cuticle fates^{4,5}.

nkd is an embryonic lethal recessive zygotic mutation that produces multiple segmentation defects, the most prominent of which is the replacement of denticles by excess naked cuticle⁶ (Fig. 1a, d). This phenotype is also seen in embryos exposed to excess Wg⁷, as well as in embryos lacking both maternal and zygotic contributions from any of three genes that antagonize Wg: *zeste-white3/glycogen synthase kinase 3 β* (*zw3/gsk3 β*) (ref. 8), *D-axin*

(ref. 9) and *D-Apc2* (ref. 10). In *nkd* embryos, *hh* and *en* transcripts initiate normally but accumulate in broad stripes including cells further from the source of Wg³, as if those cells are hypersensitive to Wg (Fig. 1e). Next, a stripe of new *wg* transcription appears just posterior to the expanded Hh/En stripe (Fig. 1f). This extra *wg* stripe requires both *wg* and *hh* activity¹¹ and is required for the excess naked cuticle seen in *nkd* mutants¹². The death of cells expressing Hh/En contributes to the marked shortening of *nkd* mutant cuticles¹³ (Fig. 1d).

To clone *nkd*, we identified a fly stock with a *lacZ* P-element insertion in the *nkd* gene. We used DNA adjacent to the P insert to

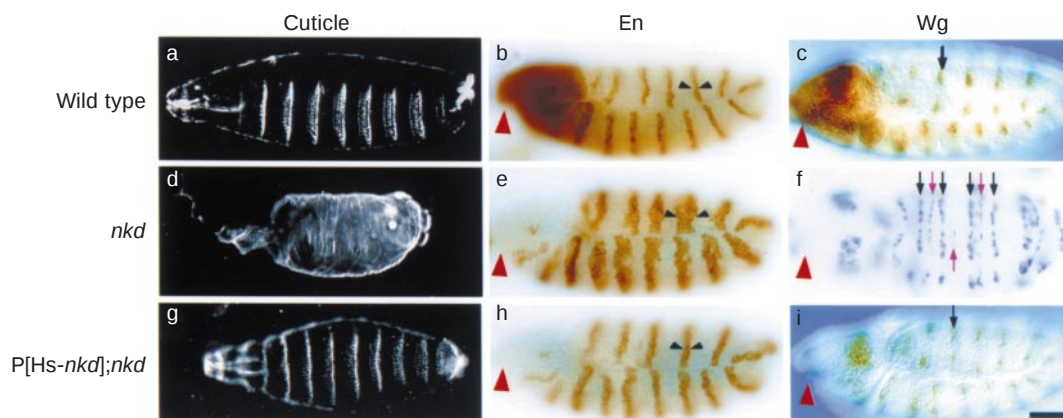


Figure 1 *nkd* embryonic phenotype and rescue by *nkd* cDNA. Wild-type ventral cuticle (a) and stage 11 embryonic En (b) and Wg (c) stains. Homozygous strong *nkd* mutants (*nkd*^{7H16} or *nkd*^{7E89}) have excess naked cuticle (d), widened En staining (e; black arrowheads) and normally spaced endogenous *wg* stripes (f; black arrows) interspersed with ectopic *wg* stripes (red arrows), most commonly in alternate parasegments. Ubiquitous Nkd expression in homozygous *nkd*^{7H16} or *nkd*^{7E89} embryos rescues the naked cuticle phenotype (g), narrows En stripes to 2–3 cells wide (h) and restores the wild-type

Wg expression pattern (i). Homozygous mutant *nkd* embryos are identified by the absence of *hunchback-lacZ* expression (red arrowheads) on the TM3 balancer chromosome. The most potent rescue was observed when 2–4 h P[HS-*nkd*]; *nkd*^{7H16}/TM3 embryos were heat shocked for 15 min at 37 °C, which resulted in 0% 'strong', 12% 'moderate', 12% 'weak' and 76% wild-type cuticles ($n = 101$) (see Methods for cuticle scoring). Scale bar, 60 μ m.

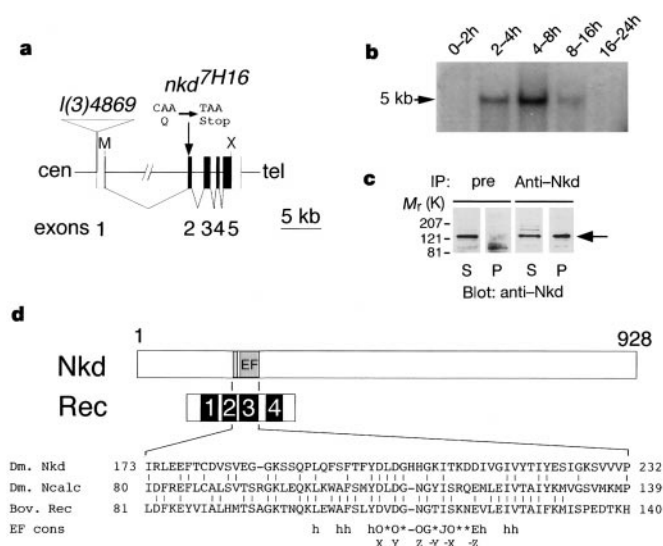


Figure 2 The *nkd* gene. **a**, Genomic map of *nkd* relative to centromere (cen) and telomere (tel) of chromosome 3L, band 75F. P-element transposon I(3)4869 is upstream of exon 1. *nkd* has five exons with a ~25-kb first intron. The putative initiator methionine (M) is near the 3' end of the first exon, whereas the stop codon (X) is in the middle of exon 5. A nonsense mutation at codon 60 (arrow above exon 2) was found in *nkd*^{7H16}. **b**, Developmental embryonic northern blot. Hours AEL shown above each lane. **c**, Immunoprecipitation of Nkd from 3–8-h embryonic extracts by anti-Nkd antibody. pre, preimmune antisera; Anti-Nkd, immune affinity purified anti-Nkd antisera; S, supernatant; P, pellet. A specific band (arrow) migrates above the 120K marker. **d**, Alignment of EF-

hand similarity between Nkd and third EF hand in recoverin family of proteins. Amino-acid identities between *Drosophila* Nkd (Dm. Nkd), *Drosophila* Neurocalcin (Dm. Ncalc; Genbank accession no. 1171668) and bovine Recoverin (Bov. Rec; Genbank accession no. 494545) are designated by vertical bars. Consensus EF-hand residues (EF cons) are shown in key symbols used by Stryer (above) or Kretsinger (below)³⁰. h, hydrophobic residue; E, acidic residue, usually glutamic acid; O, oxygen-donating residue that binds Ca²⁺; G, glycine; asterisk, variable amino acid; J, hydrophobic residue; X, Y, Z, coordinates of Ca²⁺ binding in three-dimensional space.

probe complementary DNA and genomic libraries, resulting in a genomic DNA walk of about 80 kilobases (kb) (Fig. 2a). A single 5-kb messenger RNA transcript derived from 40 kb of genomic DNA is expressed zygotically in a striped pattern (Figs 2a, b and 3). Consistent with the lack of a maternal requirement for *nkd* (S. DiNardo, personal communication), the transcript is absent from maternally derived embryonic RNA from 0–2 h AEL (Fig. 2b). The longest cDNA, 4,954 base pairs (bp), has an open reading frame (ORF) of 2,784 bp, encoding a 928-amino-acid, relatively basic (calculated isoelectric point, $pI = 9.1$) and largely hydrophilic protein. Single-stranded conformation polymorphism (SSCP) analysis and direct genomic sequencing reveals a nonsense mutation Q60stop in *nkd*^{7H16}, predicting a truncated protein of 59 amino acids (Fig. 2a). The identity of the *nkd* cDNA was further confirmed by its ability, when activated with a heat-shock promoter, to rescue the naked cuticle phenotype and the En and Wg expression abnormalities in *nkd* mutants (Fig. 1g–i).

Nkd has significant similarity to the high-affinity Ca^{2+} -binding EF hand of the recoverin family of myristoyl switch proteins (Fig. 2d; 39% amino-acid identity, 63% similarity with *Drosophila* neurocalcin, $P = 5 \times 10^{-6}$). EF hands are conserved Ca^{2+} -binding motifs that usually occur in pairs, although they have been observed singly. Mouse and human expressed sequence tag clones encoding EF-hand sequences similar to fly Nkd may be vertebrate Nkd homologues, a possibility we are currently testing. No obvious homologue has been identified in *Caenorhabditis elegans*.

Affinity-purified anti-Nkd antisera made against either of two parts of the protein detect a segmentally repeated cytoplasmic

distribution very similar to the embryonic RNA pattern (Fig. 3a, b). No staining is detected in *nkd*^{7H16} mutant embryos, and high-level ubiquitous expression is seen in heat-shocked P[Hs-*nkd*] embryos (Fig. 3b). Nkd antibody immunoprecipitates from embryonic protein extracts a protein that runs at a slightly higher relative molecular mass (M_r) than its predicted M_r of 102,000 (102K) (Fig. 2c).

nkd transcription is initiated in embryos during the late cellular blastoderm stage in broad anterior and posterior domains, in a manner reminiscent of gap genes (Fig. 3a, stage 6). During early germ-band extension (stage 8–9), *nkd* transcription is nearly ubiquitous, with higher RNA levels in the 2–3 cell rows posterior to the Hh/En stripe that require *nkd* to limit Hh/En production (Fig. 3a, stage 9). At this stage, Wg is evenly distributed on both sides of the stripe of cells that express *wg* RNA⁵. During full germ-band extension, *nkd* expression is most abundant anterior to, and lower just posterior to, the Hh/En stripe (Fig. 3a, stage 10–11). *nkd* RNA is lower still in the Hh/En-producing cells. Hh signalling in the Hh/En cells excludes Wg protein during this time, resulting in asymmetric Wg distribution with an anterior bias⁵. *nkd* mutants do not develop this anterior bias of Wg¹⁴, indicating that *nkd* may be required for *hh* to exclude Wg from Hh/En cells. Nonetheless, after embryonic stage 10, Wg protein and *nkd* RNA are found together in many tissues. *nkd* RNA and *wg* RNA are produced in overlapping patterns in imaginal discs and other larval tissues, with *nkd* domains being slightly broader than those of *wg* (Fig. 3d).

We tested whether *nkd* is regulated by Wg activity using gain- and loss-of-function experiments. In *wg* mutant embryos, *nkd* tran-

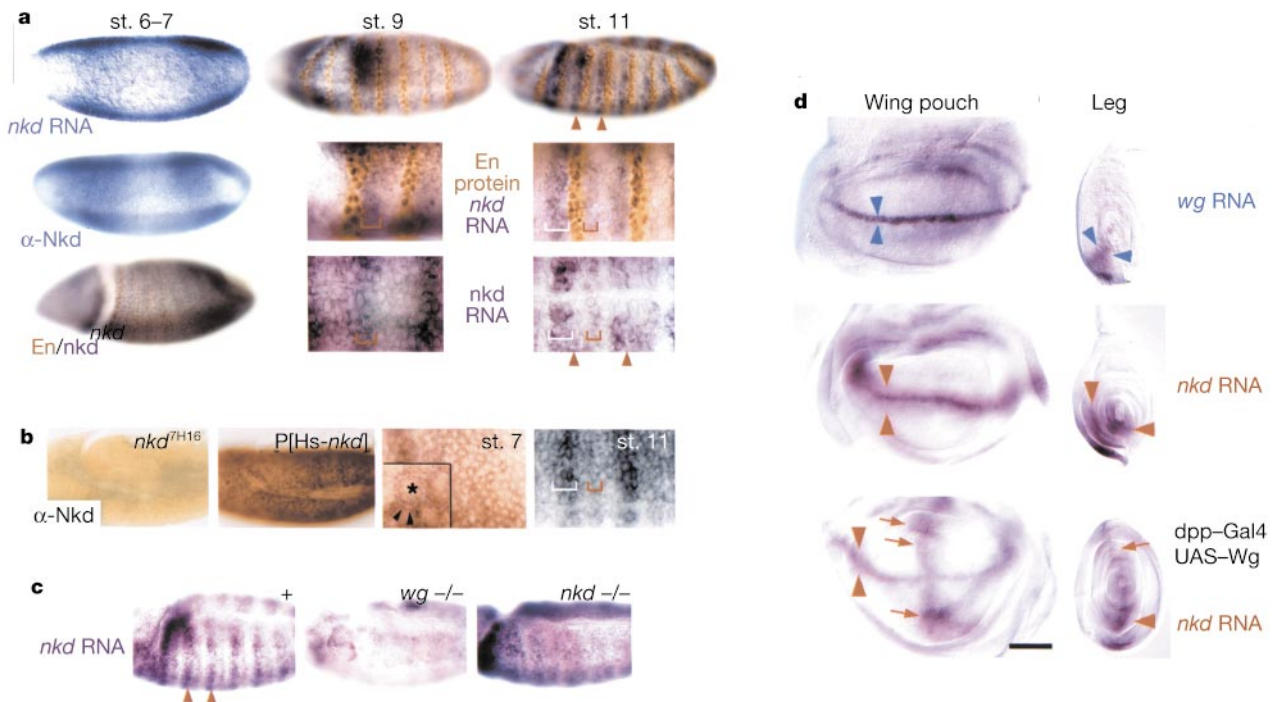


Figure 3 *nkd* expression and dependence on *wg*. **a**, *nkd* expression during embryonic stages 6–7 (left), 9 (middle) and 11 (right); anterior is left in all panels. Stage 6 *nkd* RNA (top) and Nkd protein (Anti-Nkd; middle) patterns are similar. Bottom, embryo labelled for En protein (brown) and *nkd* RNA (purple). Stages 9 and 11: *nkd* RNA with and without En protein patterns at low (top row) and higher power (bottom two rows). *nkd* accumulates posterior to the Hh/En stripe during stage 9 (red bracket). Later (stage 11), *nkd* RNA is highest in the 2–3-cell rows anterior to the Hh/En stripe (white bracket), with lower levels just posterior to the En stripe (red bracket). Parasegmental grooves are marked with red arrowheads in **a** and **c**. **b**, Anti-Nkd antiserum does not stain *nkd*^{7H16} embryos (left) and stains all cells after P[Hs-*nkd*] embryos are briefly heat pulsed at 37 °C (middle left). Wild-type stage 7 (middle right) and 11 (right) embryos stained with anti-Nkd antiserum reveal

cytoplasmic and plasma membrane-associated (inset, arrowheads) epidermal staining similar to the RNA pattern. Asterisk, nucleus; brackets as in **a**. **c**, *nkd* *in situ* hybridization to stage 11 wild-type (+), *wg* and *nkd* mutant embryos. **d**, Regulation of *nkd* expression by *wg* in third instar wing pouch (left) and leg (right) imaginal discs. Top: *wg* RNA defines the presumptive wing margin and ventral–anterior sector of the leg disc (blue arrowheads). Middle: *nkd* expression (red arrowheads) is broader than that of *wg*. Bottom: Ectopic *nkd* (red arrows) accumulates perpendicular to wing margin or in the dorsal leg disc when *dpp*-Gal4 transgene drives UAS-*wg*. Scale bar, 85 μ m in **a** left and top, **b** left two panels and **c**; 35 μ m in **a** right and bottom, **b** right two panels and **d** left; 7 μ m in **b** middle right inset; and 70 μ m in **d** right.

scription initiates normally but is markedly reduced by stage 11 (Fig. 3c). *nkd* transcript accumulates to higher levels and more broadly across the segment in *nkd* mutant embryos (Fig. 3c), presumably owing to the lack of negative feedback that Nkd protein normally provides to its own Wg-dependent expression. *nkd*

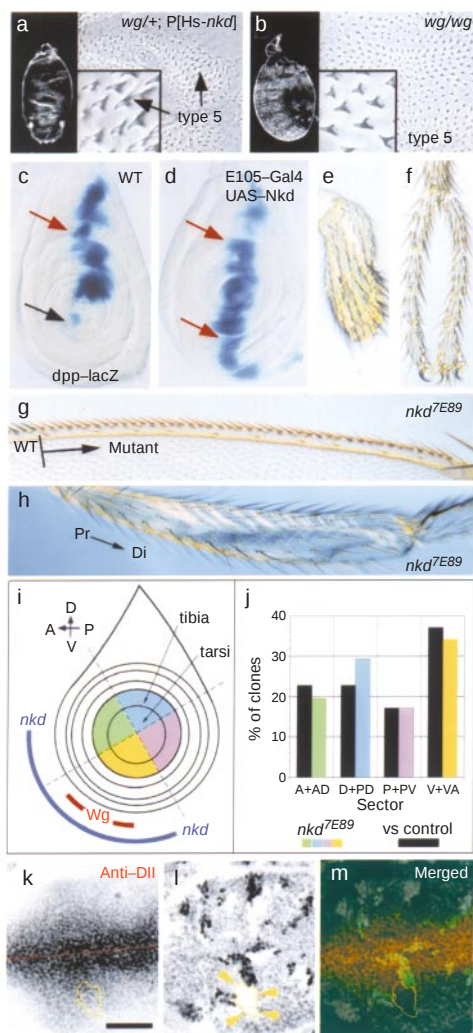


Figure 4 Consequences of altered *nkd* activity. **a**, Ubiquitous Nkd expression using P[Hs-*nkd*] in *wg*^{+/+} embryos before 4 h AEL results in shortened embryos (left, dark field) with denticle fusions and excess type 5 denticles (right, and magnified inset). **b**, *wg/wg* mutant embryos display a uniform lawn of type 5 denticles and are unaffected by excess *nkd*. **c**, X-gal-stained *dpp-lacZ* leg disc. *dpp* is expressed along the A–P axis at high levels dorsally (red arrow) and lower levels ventrally (black arrow). **d**, E105–Gal4; UAS-*nkd* imaginal discs have equally high levels of *dpp-lacZ* expression dorsally and ventrally (red arrows). **e**, E105–Gal4; UAS-*nkd* legs are variably truncated. **f**, B119–Gal4; UAS-*nkd* legs are variably duplicated. **g,h**, Phenotypically normal *nkd*^{7E89} mutant wing margin (**g**) and leg clone (**h**) marked with the bristle marker *yellow*. **i**, Leg disc fate map with sectors of unique bristle identities labelled as follows: green, anterior + anterior dorsal; blue, dorsal + posterior dorsal; lavender, posterior + posterior ventral; yellow, ventral + anterior ventral. Bristles in tibial and tarsal leg segments were scored. Perimeter of approximate domains of Wg (red) and *nkd* (blue) expression are designated with lines. **j**, Per cent of *nkd*^{7E89} clones (*n* = 39; coloured bars) and control clones (*n* = 28; black bars) as a function of leg disc quadrant as shown in **i**. Within each sector, *nkd* and control clones occur with comparable frequency. **k–m**, Negative (**k,l**) and positive (**m**) confocal images of horizontally oriented wing disc margin primordia harbouring multiple *nkd*^{6G33} clones (marked by lack of P-myc stain in **l**; green in merged image of **m**) stained with anti-Dll (**k**, red in **m**). Representative clone spanning region of greatest drop-off in Dll stain is noted by a yellow line around its perimeter, highlighted by arrows in **l**. Twin spots are marked by more intense P-myc stain throughout the disc (**l,m**). Scale bar, 110 μ m in **a** and **b**, left; 40 μ m in **c** and **d**; 120 μ m in **e–h**; 35 μ m in **k–m**.

expression is enhanced when Wg is ubiquitously expressed (not shown). Misexpression of either Wg or an activated form of the *wg*-signal transducer Armadillo (UAS-Arm^{S10})¹⁵ in wing, leg, haltere and antennal imaginal discs results in similar patterns of ectopic *nkd* transcription (Fig. 3d). Arm^{S10}-induced *nkd* transcript obeys sharp boundaries consistent with a cell-autonomous induction of *nkd* by Wg (not shown).

If loss of *nkd* mimics the effect of excess Wg, then excess Nkd should mimic loss of *wg*. When P[Hs-*nkd*] is used to overexpress *nkd* in otherwise wild-type embryos, rare cuticles with weak *wg*-like denticle fusion phenotypes are observed (not shown), similar to those seen when *zw3* is overexpressed¹⁶. Nkd is more potent in a sensitized *wg*^{+/+} background. In *wg*¹¹¹⁴/+ embryos, induction of P[Hs-*nkd*] before 4 h AEL results in decreased *en* and *wg* expression (not shown). *wg*¹¹¹⁴/+ embryos are patterned normally, but practically all *wg*¹¹¹⁴/+ embryos exposed to high levels of Nkd secrete cuticles with denticle belt fusions and an excess of the predominant denticle type made by *wg/wg* embryos (Fig. 4a, b).

Misexpressing *nkd* during larval development using UAS/Gal4 transgenes¹⁷ results in adult phenotypes that are indistinguishable from many *wg* loss-of-function phenotypes. The phenotypes that we observed include (1) wing-to-notum transformations^{18,19}; (2) leg truncations (Fig. 4e) and duplications^{18,19} (Fig. 4f); (3) loss, lateral displacement and disorientation of sternite bristles²⁰; (4) haltere loss¹⁹; (5) ventral eye reduction²¹; (6) loss of wing margin²²; (7) extra wing anterior crossveins (C. Conley and S. Blair, personal communication); and (8) loss of antennae¹⁸. The gene dosage of the Wg pathway influences the effect of ectopic Nkd: loss of one wild-type copy of *porcupine* (*porc*), *wg*, *dishevelled* (*dsh*) or *arm* enhances, and

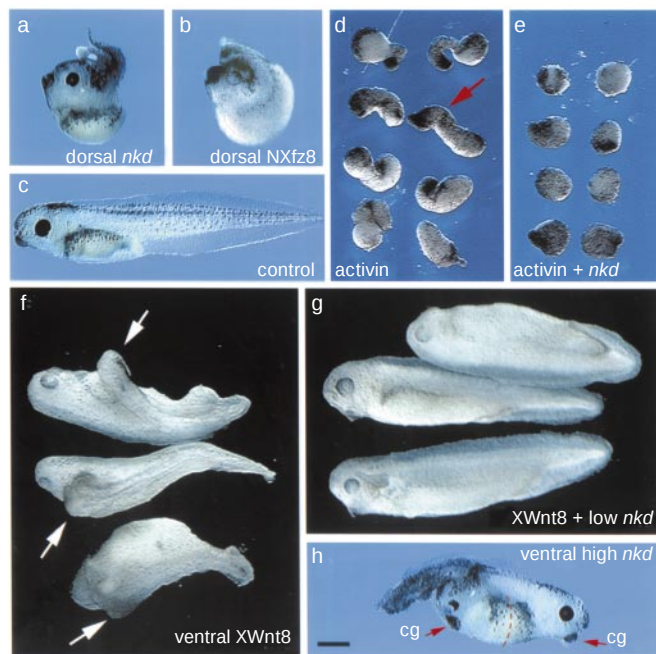


Figure 5 Misexpression of fly Nkd in *Xenopus* embryos and animal caps. Embryos with shortened A–P axes produced by injection of 2.0 ng *Drosophila nkd* RNA (**a**) or the dominant inhibitory Xfz8 (**b**) into dorsal blastomere at 4-cell stage. **c**, Injection of water has no effect. Animal caps explanted from embryos injected with *activin* RNA (**d**) elongate (red arrow) when cultured. Simultaneous expression of *activin* and *nkd* RNA (**e**) blocks elongation. *nkd* RNA injected without *activin* has no effect on animal cap elongation (not shown). **f**, Embryos receiving ventral blastomere injection of 0.5 pg Xwnt8 RNA produce partial duplicated axis (white arrow). **g**, Embryos injected with 0.5 pg Xwnt8 and 7 pg fly *nkd* do not develop ectopic axes. **h**, Ventral injection of 2 ng *nkd* RNA results in head duplication. Note the duplicated eye and cement gland (cg with arrow) and duplicated abdominal pigment (dashed line). The image in **b** has been previously published²⁵. Scale bar, 750 μ m in **a–c**, **f–h**; 375 μ m in **d** and **e**.

loss of *zw3* and *nkd* suppresses, the UAS-*nkd* overexpression phenotypes (data not shown).

During leg development, Wg and Decapentaplegic (Dpp; a signalling protein related to vertebrate bone morphogenetic proteins) act as mutually antagonistic determinants of dorsal and ventral identity²³. Dpp is expressed at high levels at the anterior-posterior (A-P) boundary dorsally, and at lower levels ventrally (Fig. 4c). Wg/Dpp juxtaposition results in leg disc eversion and outgrowth during pupal morphogenesis. Excess Nkd expressed throughout the disc results in high levels of *dpp-lacZ* expression along the entire A-P border of the disc (Fig. 4d), and scant *wg-lacZ* expression (not shown), similar to that seen when *wg* activity is reduced²³. These discs give rise to variably truncated legs (Fig. 4e), indicating that excess Nkd can antagonize the normal effects of *wg*. More restricted ventral Nkd misexpression results in duplicated legs (Fig. 4f), which may arise by an abnormal juxtaposition of cells still expressing Wg and cells in which excess Nkd results in decreased Wg, and hence derepressed *dpp*. Thus two Wg/Dpp boundaries are created and appendages are duplicated.

We induced *nkd* loss-of-function clones in imaginal discs and adult structures using two strong *nkd* alleles, *nkd*^{7H16} and *nkd*^{7E89}, and one moderately severe allele, *nkd*^{9G33}, all of which are embryonic lethal. *nkd* alleles were originally generated in the genetic background of a weak allele of the pair-rule gene *hairy* (*h*¹). *h*¹ clones give rise to ectopic wing-vein bristles and thoracic microchaetes. Furthermore, *nkd* and *h* genetically interact: *nkd*, *h*¹/*h*^{null} is lethal, whereas *h*¹/*h*^{null} is viable (A. Martinez-Arias, personal communication). Therefore we also generated clones of the strong allele *nkd*^{7E89} from which *h*¹ had been removed. In many tissues where Wg signals control pattern, including the wing (Fig. 4g), leg (Fig. 4h), thorax, abdomen, haltere and eye, we observed phenotypically normal *nkd* clones. *h*¹, *nkd*^{7H16}, but not *h*⁺, *nkd*^{7E89} or *h*¹, *nkd*^{9G33} clones give rise to a rough eye phenotype and loss of wing margin bristle phenotype (not shown), which may be due to *h*-*nkd* interactions.

To test whether *nkd* clones arise at biased locations, we scored clones in adult legs marked with the bristle marker *yellow* (Fig. 4i, j). *nkd* and control clones appeared with similar frequency in each leg quadrant (Fig. 4i). We assayed the expression of Wg target genes in *nkd* clones to look for subtle changes in gene expression that might be compatible with normal tissue patterning. Distalless (Dll) is distributed in a broad gradient centred on the margin of the wing disc (Fig. 4k). Induction of *nkd* clones results in no apparent alteration in the Dll expression gradient within, or adjacent to, multiple clones (Fig. 4k-m). No changes in cytoplasmic Arm accumulation, an indicator of Wg activity, are noted within or adjacent to *nkd* clones (data not shown).

We investigated whether fly *nkd* can alter Wnt signalling in a vertebrate by injecting mRNA into *Xenopus* embryos. Dorsal blastomere injection into 4-cell embryos of RNAs encoding the Wnt antagonist FrzB²⁴, as well as dominant inhibitory forms of Xfrizzled 8 (NXfz8)^{25,26} (Fig. 5b), Dishevelled (Xdd)²⁷ and Wnt-8 (DN-Xwnt-8)²⁸ results in marked truncations of the A-P axis. Injections of fly *nkd* RNA result in very similar effects (Fig. 5a). Injection of 0.5 ng *nkd* RNA caused severe A-P truncations in 52% (*n* = 19) of the embryos, with milder effects in the rest, whereas 2.0 ng *nkd* RNA increased the penetrance to 97% (*n* = 38), indicating a dose-response relationship. Antagonism of Wnt function apparently blocks cell movements that drive the elongation of the gastrula and neurula^{25,27}. These movements drive the elongation of ectodermal explants (animal caps) induced to form mesoderm by activin (Fig. 5d). Wnt antagonists (such as NXfz8 or Xdd) block elongation of explants without inhibiting mesoderm induction^{25,27}. Injection of 2.0 ng *nkd* RNA into animal caps mimics this inhibitory effect, blocking elongation in response to activin (2.5 pg mRNA) in 90% of explants (*n* = 47) (Fig. 5e).

Ectopic ventral expression of Wnts before the onset of zygotic transcription results in duplication of the dorsal axis, which can be

blocked by Wnt antagonists such as NXfz8 (ref. 25), Xdd²⁷ or FrzB²⁴. *nkd* also blocks Wnt-mediated axis duplication. Ventral blastomere injection of 0.5 pg XWnt8 RNA resulted in ectopic dorsal axes in 56% of embryos (*n* = 88) (Fig. 5f), about half of which formed complete anterior structures. Co-injection of 3.5 pg *nkd* RNA reduced the frequency of ectopic axes to 42% (*n* = 72), with only 10% of embryos forming complete anterior structures. Co-injection of 35 pg *nkd* RNA gave only 19% (*n* = 80) partial secondary axes, with no complete axes. At these doses, *nkd* RNA alone has no discernible effect on development when injected into either dorsal or ventral blastomeres. Higher doses (350 pg–2 ng) of *nkd* RNA injected alone into ventral blastomeres give rise to ectopic heads (Fig. 5h), complete with eyes and cement glands. This striking phenotype has been attributed to antagonism of Wnt activity^{26,29}.

Our data show that *nkd* antagonizes Wg/Wnt signalling. Does *nkd* affect Wnt synthesis or transport, or determine how cells respond to Wnt? Because Wg and other Wnts are autoregulatory in many contexts, Nkd could affect the quantity or distribution of Wg either directly by controlling Wg synthesis or transport, or indirectly by reducing the positive feedback of Wg activity on its own synthesis. Unfortunately, the lack of an *nkd* clonal phenotype in the fly precludes definitive determination of cell-autonomous function. However, the first known gene expression defect in *nkd* mutants is in cells distant to Wg producers, indicating that *nkd* may act first in Wg-receiving cells. In addition, *nkd* RNA injected into *Xenopus* embryos does not alter the accumulation of epitope-tagged XWnt8 (not shown), indicating that *nkd* may block the response to XWnt8.

Inducible antagonists limit the effective duration, range of action or activity of a signal, and can act cell autonomously or non-autonomously¹. Near-saturating genetic screens have revealed that *nkd* and *patched* (*ptc*; an inducible antagonist for Hh signalling) are probably the only *Drosophila* Wg or Hh pathway genes, other than Wg or Hh themselves, whose expression and genetic requirement are exclusively zygotic. All other known components of both signalling pathways are maternally provided. Evolutionary selective pressure apparently resists duplications of zygotically active inducible antagonist genes. Antagonist gene dosage must be carefully regulated in flies and vertebrates to balance the effects of the signals. In *Drosophila*, *nkd* and *ptc* mutations have haplo-insufficient effects on cuticle pattern in combination with each other and with other segment-polarity mutants¹¹. Altered regulation of both Wnt and Hh signalling in mice and humans is implicated in precancerous and cancerous cell growth. Just as vertebrate *ptc1* regulates cell fates and is a key tumour-suppressor gene, vertebrate Nkd-like proteins may be essential for restraining Wnt activity during development and possibly cancer progression. □

Methods

Cloning of *nkd*

nkd^{7H16} and *nkd*^{7E89} fail to complement the lethality of enhancer trap line l(3)/4869. Homozygous l(3)/4869 embryos have features of weak *nkd* alleles (see below). Precise excision of the l(3)/4869 P element by transposase eliminated lethality and the weak *nkd* mutant phenotype. A 4-kb *Xba*I-*Xba*I genomic DNA fragment immediately 3' to the P-element insertion site in l(3)/4869 was cloned by plasmid rescue and used to initiate the isolation of 80 kb of genomic DNA.

SSCP analysis

We performed SSCP analysis using MDE high-resolution gel from AT Biochem. 150 ng of genomic DNA was used as the template. Primers 100–700 bp apart were used for SSCP-PCR. The sequences of the two primers that detect a nonsense mutation at codon 60 in *nkd*^{7H16} are 5'-GCTGCTGGTCAGCGAACG-3' (CTP16) and 5'-TGATGAGACTGCTGC TTAC-3' (CBP17).

Fly stocks and P-element-mediated transformation

*wg*¹¹¹⁴ was used at non-permissive temperatures, mimicking a null *wg* allele, in Figs 3c and 4a,b. P-element-mediated transformation was performed on *yw* flies, using P Δ 2-3 as a source of transposase.

Heat-shock mediated Nkd overexpression and *nkd* rescue

We made P[*Hs-nkd*] by cloning the *Kpn-Kpn* fragment from *nkd* cDNA clone C5 into *Kpn*-cut pABAL, which harbours the hsp70 promoter and 3' untranslated regions. The P[*Hs-nkd*] construct that was used for rescues lacks 5' and 3' untranslated regions and has a carboxy-terminal 9-amino-acid human *c-myc* tag. Two independent transformant lines gave quantitatively similar results in Nkd overexpression experiments in both *nkd* mutant and *wg*⁺ backgrounds. Most experiments used *nkd*^{7H16}; *nkd*^{7E89} was rescued to a similar degree by three independent insertions of P[*Hs-nkd*]. All known *nkd* alleles are embryonic lethal, and their cuticle phenotype severity can be scored as follows. 'Strong' *nkd* mutants (such as *nkd*^{7H16} and *nkd*^{7E89}) secrete cuticles with a fully exteriorized head skeleton, widely split filzkörper (posterior spiracles) and residual denticle belts only in A3 and/or A5 (or none at all), and are typically less than 75% of wild-type length. 'Weak' *nkd* cuticles (such as l(3)4869 and *nkd*^{42H1}) (W.Z. and M.P.S., unpublished data) have patchy cuticle loss, but possess largely normal filzkörper and head skeleton, and are almost wild-type length. 'Moderate' cuticles (such as *nkd*^{9G33} and *nkd*^{9H52}) have a phenotype between weak and strong.

Nkd antibody production

Recombinant Nkd protein was made as a TrpE–Nkd fusion. The fusion protein was purified from BL21 pLysS *Escherichia coli* lysates as inclusion bodies, cut from an SDS–polyacrylamide gel and injected with adjuvant into rabbits (Josman Labs). We made GST–Nkd fusion proteins to affinity purify anti-Nkd antisera by cloning similar fragments of the *nkd* cDNA into the vector pGEX-4T (Pharmacia). GST–Nkd fusion proteins were purified from BL21 pLysS *E. coli* using glutathione agarose beads coupled to an Aminolink Plus chromatography column (Pierce).

Immunocytochemistry

Embryos were washed and dechorionated in 50% bleach for 5 min, heat fixed for 10 s in a 90 °C solution of 100 mM NaCl and 0.05% Triton-X, and devitellinized in equal volumes of heptane and methanol. Standard 4% formaldehyde fixation methods failed to reveal any specific staining pattern using both Nkd antisera. P[*Hs-nkd*] embryos were heat shocked at 37 °C for 30 min and allowed to recover for 30 min before fixation. Antibodies used: Nkd rabbit polyclonal antisera (1:100); biotinylated anti-rabbit secondary antibody (1:200). Staining was with biotinylated horseradish peroxidase (HRP)/avidin (Vectastain Elite ABC, Vector Laboratories) and 3,3'-diaminobenzidine as a substrate. For the Nkd stains in Fig. 3a, b, we added 4 µl of a 6% solution of NiCl₂ to the staining reaction. Polyclonal antibodies made against two different parts of the Nkd protein (*Sma1*–*Sma1* fragment, amino acids 100–371; *Ase1*–end, amino acids 679–928) give similar embryonic staining patterns. En monoclonal antibody supernatant was used at 1:1 on embryos fixed in 4% paraformaldehyde.

Immunoprecipitation

A 3–8-h S-100 embryo extract was prepared in NP-40 buffer with proteinase inhibitor cocktail. 60 µl of extract was incubated with 1 µl of preimmune serum or 5 µl of Nkd rabbit polyclonal antiserum. The antibodies were collected with protein A–Sepharose beads, eluted and separated on an 8% SDS–polyacrylamide gel. Antibodies used: Nkd rabbit polyclonal antiserum (1:20); HRP-conjugated secondary antibody (1:20,000; Jackson ImmunoResearch Labs). HRP activity was detected with Super Signal Chemiluminescent Substrate (Pierce).

In situ hybridization

Whole-mount *in situ* hybridization was performed with digoxigenin-labelled antisense RNA (Boehringer) using the 2.7-kb *nkd* coding region as a template. The probe was gently carbonated to produce fragments that were visualized on a gel in the 300-bp range.

UAS/Gal4-mediated *nkd* overexpression

We cloned native and myc-tagged versions of Nkd into pUAS-T¹⁷ and transformed them into flies. Constructs give similar phenotypes regardless of chromosome insertion site or Gal4 line. The dosage sensitivity of Wg pathway components to Nkd overexpression was tested using the alleles *porc*^{JB}, *wg*^{IL114}, *wg*^{CK4}, *dsh*⁴⁷⁷, *dsh*^{IL26}, *zw3*^{M11}, *nkd*^{7H16}, *nkd*^{7E89} and *arm*^{YD35}. We used B119–Gal4 and E132–Gal4 to drive UAS–*nkd* in these experiments, and three phenotypes were assayed for each Gal4 line in 50–1,000 adults for each experiment. Detailed results are available upon request.

Clonal analysis

Balanced *nkd*^{7H16} and *nkd*^{9G33}, originally generated on *h*¹, *ru*, *cu*, *ca* chromosome, and *st*, *nkd*^{7E89} which is *h*¹, were mated to *w*; *FRT 80B*, and nonbalancer progeny were mated to *w*; *TM3/TM6* on food containing 50 µg ml⁻¹ Gentacin (G418) at 26 °C. Putative *nkd*, *FRT80B/TM6* recombinants were mated with different *nkd/TM3* alleles to test for complementation. Clones were induced between 3 and 48 h AEL by mating *w* or +; *nkd FRT 80B* males with *y*, *w*, *hs-FLP*; *Z*, *FRT 80B* females, where *Z* = P[*y*+, *w*+], for marking mutant bristles with yellow and mutant ommatidia with white, or *Z* = P[*hs*–*πmyc*] or P[*Ubi*–*GFP*] for marking mutant clones with anti-Myc or GFP. Progeny were heat shocked at 37 °C for 30 min–2 h to induce expression of FLP recombinase, and males were examined for the presence of mutant yellow bristles. For imaginal tissues, we used rabbit anti-Dll at 1:200; mouse anti-Myc (9E10) at 1:1; and mouse anti-Arm (N2-D71) at 1:5. Clones were detected using either heat-shock-mediated induction of *πmyc* or constitutive ubiquitous expression of GFP. For *h*¹, *nkd* clones, twin spots tended to be larger than clones, suggesting a cell-autonomous lethal effect.

Xenopus injections

The Nkd ORF with the C-terminal Myc tag was subcloned into the expression vector pCS2 (D. Turner and R. Rupp) and mRNA was prepared by *in vitro* transcription. We injected *nkd* mRNA with or without *Xwnt8* mRNA into the animal pole of one dorsal or ventral blastomere at the 4-cell stage (10 nl per cell) and cultured embryos until the tadpole stage for scoring. For mesoderm induction assays, 10 nl *nkd* mRNA (0.2 ng nl⁻¹) and/or *activin* (0.25 pg nl⁻¹) mRNA was injected into the animal pole of fertilized eggs. At the blastula stage, animal caps were explanted, cultured until siblings reached the neurula stage and then scored for elongation.

Received 5 October; accepted 6 December 1999.

- Perrimon, N. & McMahon, A. P. Negative feedback mechanisms and their roles during pattern formation *Cell* **97**, 13–16 (1999).
- Cadigan, K. M. & Nusse, R. Wnt signaling: a common theme in animal development *Genes Dev.* **11**, 3286–3305 (1997).
- DiNardo, S., Heemskerck, J., Dougan, S. & O'Farrell, P. H. The making of a maggot: patterning the *Drosophila* embryonic epidermis *Curr. Opin. Genet. Dev.* **4**, 529–534 (1994).
- Bejsovec, A. & Martinez Arias, A. Roles of *wingless* in patterning the larval epidermis of *Drosophila* *Development* **113**, 471–485 (1991).
- Sanson, B., Alexandre, C., Fascetti, N. & Vincent, J. P. Engrailed and hedgehog make the range of *Wingless* asymmetric in *Drosophila* embryos *Cell* **98**, 207–216 (1999).
- Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C. & Kluding, H. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. II. Zygotic loci on the third chromosome *Wilhelm Roux Arch. Dev. Biol.* **193**, 283–295 (1984).
- Noordermeer, J., Johnston, P., Rijsewijk, F., Nusse, R. & Lawrence, P. A. The consequences of ubiquitous expression of the *wingless* gene in the *Drosophila* embryo *Development* **116**, 711–719 (1992).
- Siegrfried, E., Chou, T. B. & Perrimon, N. *wingless* signaling acts through *zeste-white 3*, the *Drosophila* homolog of glycogen synthase kinase-3, to regulate engrailed and establish cell fate *Cell* **71**, 1167–1179 (1992).
- Hamada, F. *et al.* Negative regulation of *Wingless* signaling by D-axin, a *Drosophila* homolog of axin *Science* **283**, 1739–1742 (1999).
- McCartney, B. M. *et al.* *Drosophila* APC2 is a cytoskeletonally-associated protein that regulates *wingless* signaling in the embryonic epidermis *J. Cell Biol.* **146**, 1303–1318 (1999).
- Bejsovec, A. & Wieschaus, E. Segment polarity gene interactions modulate epidermal patterning in *Drosophila* embryos *Development* **119**, 501–517 (1993).
- Dougan, S. & DiNardo, S. *Drosophila* *wingless* generates cell type diversity among engrailed expressing cells *Nature* **360**, 347–350 (1992).
- Pazdera, T. M., Janardhan, P. & Minden, J. S. Patterned epidermal cell death in wild-type and segment polarity mutant *Drosophila* embryos *Development* **125**, 3427–3436 (1998).
- Moline, M. M., Southern, C. & Bejsovec, A. Directionality of *Wingless* protein transport influences epidermal patterning in the *Drosophila* embryo *Development* **126**, 4375–4384 (1999).
- Pai, L. M., Orsulic, S., Bejsovec, A. & Peifer, M. Negative regulation of Armadillo, a *Wingless* effector in *Drosophila* *Development* **124**, 2255–2266 (1997).
- Steitz, M. C., Wickenheiser, J. K. & Siegrfried, E. Overexpression of *zeste white 3* blocks *wingless* signaling in the *Drosophila* embryonic midgut *Dev. Biol.* **197**, 218–233 (1998).
- Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes *Development* **118**, 401–415 (1993).
- Baker, N. E. Transcription of the segment-polarity gene *wingless* in the imaginal discs of *Drosophila*, and the phenotype of a pupal-lethal *wingless* mutation *Development* **102**, 489–497 (1988).
- Couso, J. P., Bate, M. & Martinez-Arias, A. A *wingless*-dependent polar coordinate system in *Drosophila* imaginal discs *Science* **259**, 484–489 (1993).
- Shirras, A. D. & Couso, J. P. Cell fates in the adult abdomen of *Drosophila* are determined by *wingless* during pupal development *Dev. Biol.* **175**, 24–36 (1996).
- Ma, C. & Moses, K. *Wingless* and *patched* are negative regulators of the morphogenetic furrow and can affect tissue polarity in the developing *Drosophila* compound eye *Development* **121**, 2279–2289 (1995).
- Couso, J. P., Bishop, S. A. & Martinez Arias, A. The *wingless* signalling pathway and the patterning of the wing margin in *Drosophila* *Development* **120**, 621–636 (1994).
- Brook, W. J. & Cohen, S. M. Antagonistic interactions between *wingless* and *decapentaplegic* responsible for dorsal-ventral pattern in the *Drosophila* leg *Science* **273**, 1373–1377 (1996).
- Moon, R. T., Brown, J. D., Yang-Snyder, J. A. & Miller, J. R. Structurally related receptors and antagonists compete for secreted Wnt ligands *Cell* **88**, 725–728 (1997).
- Deardorff, M. A., Tan, C., Conrad, L. J. & Klein, P. S. Frizzled-8 is expressed in the Spemann organizer and plays a role in early morphogenesis *Development* **125**, 2687–2700 (1998).
- Itoh, K. & Sokol, S. Y. Axis determination by inhibition of Wnt signaling in *Xenopus* *Genes Dev.* **13**, 2328–2336 (1999).
- Sokol, S. Y. Analysis of Dishevelled signalling pathways during *Xenopus* development *Curr. Biol.* **6**, 1456–1467 (1996).
- Hoppler, S., Brown, J. D. & Moon, R. T. Expression of a dominant-negative Wnt blocks induction of MyoD in *Xenopus* embryos *Genes Dev.* **10**, 2805–2817 (1996).
- Glinka, A., Wu, W., Onichtchouk, D., Blumenstock, C. & Niehrs, C. Head induction by simultaneous repression of Bmp and Wnt signalling in *Xenopus* *Nature* **389**, 517–519 (1997).
- Flaherty, K. M., Zozulya, S., Stryer, L. & McKay, D. B. Three-dimensional structure of recoverin, a calcium sensor in vision *Cell* **75**, 709–716 (1993).

Acknowledgements

We thank L. Mathies for identifying the P element in the *nkd* gene; D. Bilder, K. Cadigan, S. Carroll, C. Goodman, A. Hudson, N. Ito, C. Kirkpatrick, A. Martinez-Arias, R. Nusse, C. Nüsslein-Volhard, G. Panganaban, N. Patel, M. Peifer, N. Perrimon, G. Rubin and A. Spradling for mutants, reagents and advice; J. Axelrod, L. Luo, R. Nusse and R. Rousset for comments on the manuscript; S. Blair, S. Conley, S. DiNardo, A. Martinez-Arias, S. Sokol and M. Deardorff for providing information before publication; M. Fish for

injection of UAS-Nkd constructs; and K. Cadigan and D. Bilder for stained material. W.Z. and J.M. were supported by postdoctoral training grants from the NIH and by the Howard Hughes Medical Institute (HHMI). K.A.W. was supported by an HHMI postdoctoral fellowship for physicians and a NIH K-08 Award. P.K. and M.P.S. are Investigators of the HHMI.

Correspondence and requests for materials should be addressed to M.P.S. (e-mail: scott@cmgm.stanford.edu). The *Drosophila nkd* cDNA has been deposited with GenBank (accession no. AF213376).

Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase

Shin-ichiro Imai, Christopher M. Armstrong, Matt Kaerberlein & Leonard Guarente

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Yeast Sir2 is a heterochromatin component that silences transcription at silent mating loci¹, telomeres² and the ribosomal DNA^{3,4}, and that also suppresses recombination in the rDNA⁵ and extends replicative life span⁶. Mutational studies indicate that lysine 16 in the amino-terminal tail of histone H4 and lysines 9, 14 and 18 in H3 are critically important in silencing, whereas lysines 5, 8 and 12 of H4 have more redundant functions^{7–9}. Lysines 9 and 14 of histone H3 and lysines 5, 8 and 16 of H4 are acetylated in active chromatin and hypoacetylated in silenced chromatin, and overexpression of Sir2 promotes global deacetylation of histones^{9,10}, indicating that Sir2 may be a histone deacetylase. Deacetylation of lysine 16 of H4 is necessary for binding the silencing protein, Sir3 (ref. 8). Here we show that yeast and mouse Sir2 proteins are nicotinamide adenine dinucleotide (NAD)-dependent histone deacetylases, which deacetylate lysines 9 and 14 of H3 and specifically lysine 16 of H4. Our analysis of two *SIR2* mutations supports the idea that this deacetylase activity accounts for silencing, recombination suppression and extension of life span *in vivo*. These findings provide a molecular framework of NAD-dependent histone deacetylation that connects metabolism, genomic silencing and ageing in yeast and, perhaps, in higher eukaryotes.

Sir2 is a limiting component that promotes longevity in yeast mother cells. Cells lacking Sir2 have a reduced replicative life span and cells with an extra copy of *SIR2* display a much longer life span than wild type⁶. This extension probably results from a hypersilencing in the rDNA which reduces recombination and the production of extrachromosomal rDNA circles, a known cause of senescence in ageing mother cells¹¹. *SIR2* homologues have been identified in many organisms ranging from bacteria to humans¹². The *Salmonella* homologue, *cobB*, has been implicated in a pyrimidine transfer reaction¹³ and both CobB and eukaryotic Sir2 proteins possess ADP-ribosyltransferase activity^{14,15}.

Because Sir2 proteins use NAD as a substrate in a ADP-ribosylation reaction, we examined whether NAD could be a co-factor necessary for deacetylase activity. We used purified recombinant Sir2 in a reaction with NAD and a peptide of the histone H3 N-terminal tail (residues 1–20) di-acetylated at lysines 9 and 14. We reacted 5 µg of recombinant yeast Sir2 (79 pmoles) (Fig. 1a), 10 µg of the H3 peptide (4.2 nmoles) and increasing concentrations of NAD and analysed the products by high-pressure liquid chromatography (HPLC). The peptide that reacted in the absence of NAD gave

rise to two peaks (3 and 5), which were analysed by mass spectroscopy (Fig. 1b; and Supplementary Information) and correspond to a monomeric (relative molecular mass (M_r) 2370) and a dimeric (M_r 4740) peptide, the latter probably due to oxidation of the peptide at the carboxyl cysteine residue. The same species were observed in reactions with a control bacterial preparation ('pET' in Fig. 1a) in the presence of NAD.

The addition of NAD to the reaction containing Sir2 gave rise to three additional peaks (1, 2 and 4) and an alteration in peak 3, which were also analysed by mass spectroscopy (Fig. 1c–f; and Supplementary Information). These peaks did not correspond to ADP-ribosylated species, but, rather, to deacetylated species of peptide (Fig. 1g). Peak 4 corresponded to the singly deacetylated dimer (M_r 4698); peak 3 also contained the doubly deacetylated dimer (M_r 4656); peak 2 corresponded to the triply deacetylated dimer (M_r 4614), and peak 1 to the singly deacetylated monomer (M_r 2328). We estimate that at least 27% of the input peptide was deacetylated

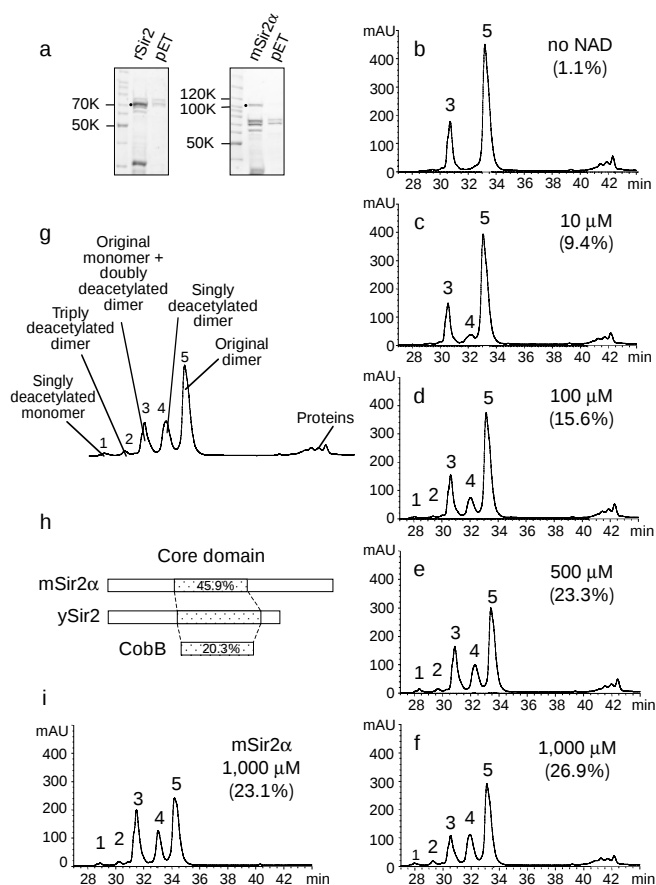


Figure 1 *In vitro* deacetylation assays of the H3 peptide (residues 1–20) di-acetylated at lysines 9 and 14 by recombinant yeast Sir2. **a**, Coomassie blue-stained gel after SDS-PAGE of purified recombinant yeast (rSir2) and mouse Sir2 (mSir2α) proteins. Full-length proteins are indicated by dots. Vector controls (pET) were prepared analogously to recombinant proteins. **b–f**, HPLC chromatograms showing absorbance at 220 nm of products of deacetylation assays with yeast Sir2 and the indicated concentrations of NAD. The efficiencies of the reactions are calculated as a fraction of the areas under peaks 1, 2 and 4, compared with the area under all of the peaks. **g**, Contents of peaks 1–5. From the relative abundances of peaks, the dimer peptide is probably a better substrate than the monomer for Sir2. **h**, Comparison of mSir2α, yeast Sir2 (ySir2) and CobB. Core domains are shown as shaded boxes with the percentages of amino-acid identity to the ySir2 core domain. **i**, The HPLC chromatogram of the product of deacetylation assay with 10 µg of recombinant mSir2α protein at 1 mM NAD. The calculated efficiency of the reaction is indicated.

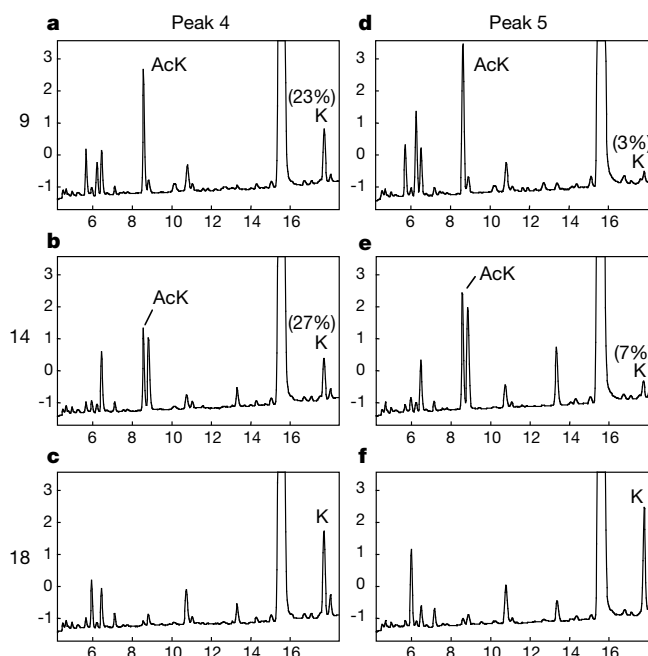


Figure 2 Amino-terminal sequencing of peaks 4 and 5 of the Sir2 deacetylase reaction at 1 mM NAD as determined by Edman degradation. Chromatograms at positions 9 (**a,d**), 14 (**b,e**) and 18 (**c,f**) are shown. In peak 4, about 23% of Lys 9 (**a**) and 27% of Lys 14 (**b**) are deacetylated. In peak 5, Lys 9 and Lys 14 are essentially fully acetylated (**d,e**).

Unacetylated Lys 18 of peaks 4 and 5 is shown for comparison (**c,f**). The peak to the right of the acetylated lysine at position 14 corresponds to the chromatographic position of alanine, and represents a preview of Ala 15 in cycle 14.

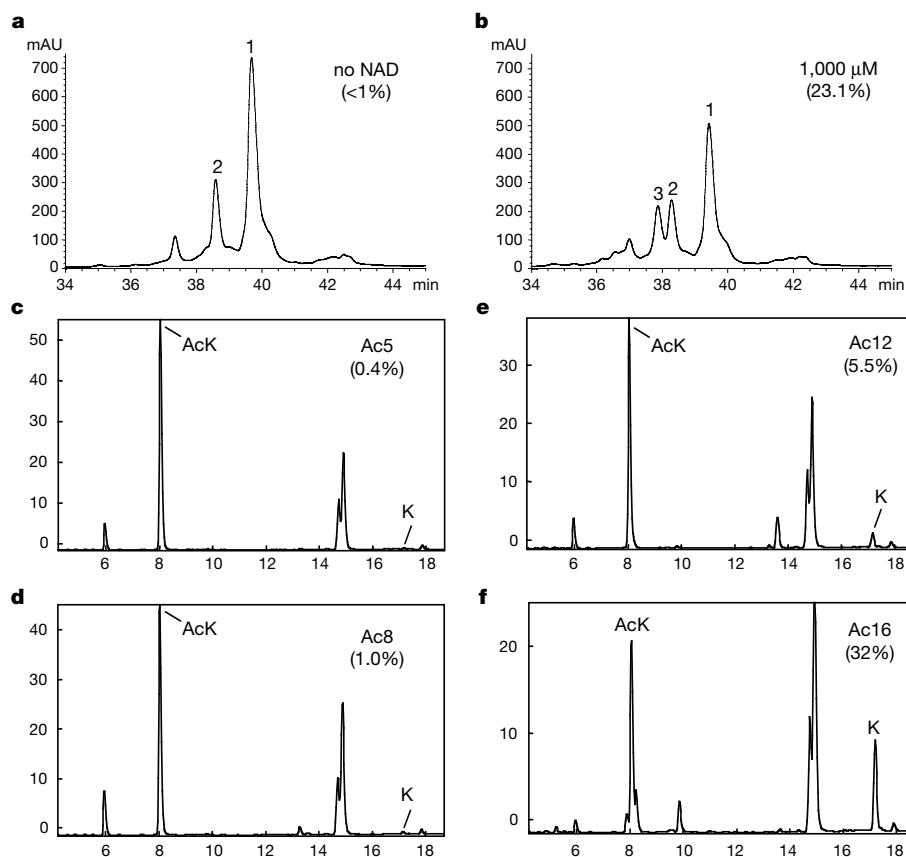


Figure 3 The deacetylation activity of yeast Sir2 on the H4 peptide (residues 2–19) tetra-acetylated at lysines 5, 8, 12 and 16. **a,b**, HPLC chromatograms of products of deacetylation assays with Sir2 and the indicated concentrations of NAD. The efficiencies of the reactions are calculated from the areas under peaks. Peaks 1 and 2 correspond to the dimeric and monomeric forms, respectively, of the input peptide. Peak 3 corresponds to the deacetylated product. **c–f**, Amino-terminal sequencing of peak 3. Chromatograms

of positions 5, 8, 12 and 16 are shown. Substantial deacetylation (32%) is observed for only Lys16. The sum of deacetylation of all four residues, about 40%, is less than the theoretical 50% expected for the singly deacetylated dimeric peptide, most probably owing to contamination of peak 3 by peak 2. As a control, peak 1, corresponding to the original dimer peptide, shows minimal deacetylation at positions 5, 8, 12 and 16 (0.1%, 0.1%, 0.7% and 0.5%, respectively).

by Sir2. The concentration of NAD that allowed the reaction to proceed to ~50% of the maximal level was ~100 μ M.

To analyse these reaction products further, we subjected peak 4, the singly deacetylated dimer, along with the input peak 5, to N-terminal protein sequencing by Edman degradation (Fig. 2). The only differences between the input and reacted peaks occurred at lysines 9 and 14; 23–27% of the acetyl lysines at each position were deacetylated by Sir2 in the presence of NAD. Lysine 18 of peaks 4 and 5 was unacetylated. NADH, NADP and NADPH did not promote a significant level of deacetylation by Sir2, and neither NADH nor NADP inhibited the activity of NAD in this reaction (see Supplementary Information). Thus, Sir2 is an NAD-dependent histone deacetylase that can deacetylate either Lys 9 or Lys 14 of the H3 N-terminal tail. The number of moles of peptide deacetylated greatly exceeds that of Sir2, indicating that Sir2 catalyses multiple reaction cycles.

We then examined the Sir2 deacetylase activity using a 20-residue peptide of the N terminus of histone H4 completely acetylated at lysines 5, 8, 12 and 16 and containing a carboxyl cysteine. In the absence of NAD, this peptide gave rise to two prominent peaks (1 and 2) in HPLC corresponding to the dimeric and monomeric peptides, respectively (Fig. 3a). Addition of NAD elicited a third prominent peak (3, Fig. 3b) corresponding to a singly deacetylated species, which was collected and analysed by Edman degradation. Deacetylation selectively occurred at Lys 16 (Fig. 3c–f). About 32% of the Lys 16 was deacetylated as opposed to less than 6% of the other lysines. This specificity of Sir2 corresponds to the primary importance of Lys 16 of histone H4 in silencing.

We tested the effect of a potent inhibitor of histone deacetylases, trichostatin A (TSA)¹⁶. TSA was totally incapable of inhibiting deacetylation by Sir2 at 400 nM (Fig. 4a, b), a concentration that

inhibits histone deacetylases, including Rpd3 family members¹⁷. Consistent with its reported ADP-ribosyltransferase activity, Sir2 transferred ³²P from NAD to intact histone H3 (Fig. 4c) or to the di-acetylated H3 peptide (Fig. 4d). Transfer of label to the peptide was assayed using thin-layer chromatography (TLC), which revealed not only this transfer but also a substantial amount of NAD hydrolysis. We then tested a known inhibitor of mono-ADP ribosyltransferases, coumermycin A1 (ref. 18). Both the ADP ribosylation (Fig. 4c, d) and NAD hydrolysis (Fig. 4d) were inhibited by this drug at 200 μ M, a concentration that inhibits other mono-ADP-ribosyltransferases. Coumermycin A1 was not capable of inhibiting deacetylation of the H3 tail by Sir2 (Fig. 4e). Thus, the proposed activity of ADP-ribosyltransferase (and the NAD hydrolase) is fundamentally distinct from this NAD-dependent deacetylation activity.

We characterized a mouse *SIR2* homologue called *mSir2 α* (GenBank accession number AF214646). The conserved region of this protein resembles yeast Sir2 more closely than do other mouse Sir2 proteins (Fig. 1h; and Supplementary Information). To determine whether the mouse Sir2 homologue *mSir2 α* would catalyse this deacetylation reaction, we incubated purified recombinant *mSir2 α* with the di-acetylated H3 peptide and analysed the reaction products by HPLC (Fig. 1a, i). The murine protein gave rise to the same array of products with a similar yield to that of the yeast enzyme, suggesting that mammalian Sir2 proteins are also mediators of transcriptional silencing.

To determine whether the histone deacetylase activity of Sir2 is required for *in vivo* functions, two alanine mutations in highly conserved residues of the core domain (Gly 270 and Asn 345) were introduced by site-directed mutagenesis (Fig. 5Aa). Six-histidine-tagged proteins were purified from *Escherichia coli* (Fig. 5Ab) and assayed for deacetylase activity with the di-acetylated H3 peptide (Fig. 5Ac–f). Mutant 345 was inactive, providing strong evidence that the deacetylase activity is an inherent property of Sir2, whereas mutant 270 showed a high level of activity—about 80% of that of wild type. These proteins were also analysed for ADP-ribosyltransferase activity using the intact histone H3 as substrate (Fig. 5Ag). Mutant 345 was inactive and mutant 270 now displayed only a very weak activity, about 7% of that of wild type.

To examine the functions of these mutants *in vivo*, we used a yeast indicator strain in which the endogenous *SIR2* had been deleted and the wild-type or mutant *SIR2* genes were integrated back into the genome. The mutant proteins were stably expressed (Fig. 5Ba). Silencing of the *HML α* locus was assayed by mating with a haploid strain of opposite mating type and monitoring the appearance of diploids on selective media (Fig. 5Bb). Mutant 270 was mating proficient, which indicates that silencing was intact, and mutant 345 was defective. Telomere silencing was determined by repression of the telomere-positioned *URA3* gene on media containing 5-fluoro-orotic acid (FOA), and this assay showed that mutant 270 had partial silencing and 345 was defective (Fig. 5Bc). Silencing in the rDNA was determined by repression of rDNA-positioned *ADE2* on adenine-free media, showing that mutant 270 silenced as well as the wild type, and that mutant 345 was defective (Fig. 5Bd).

Recombination in the rDNA was assayed by loss of *ADE2* on media limiting in adenine, giving rise to a red pigment. Half-red/white sector colonies indicate *ADE2* loss in the first generation after plating, and the frequency of these colonies compared with *Ade*⁺ colonies is a direct measure of the recombination rate in the rDNA. Wild-type *SIR2* suppressed recombination about 12-fold compared with the *sir2* deletion strain; mutant 270 showed a high degree of suppression in the range of wild type; and mutant 345 was as defective as the *sir2* deletion (Fig. 5Be). Replicative life spans were also determined as another measure of *SIR2* function in the rDNA (Fig. 5Bf). The *sir2* deletion shortened life span about 50% as compared with wild type, as expected. Mutant 270 complemented the *sir2* deletion to give a wild-type life span and mutant 345 was completely defective in complementing the life span defect.

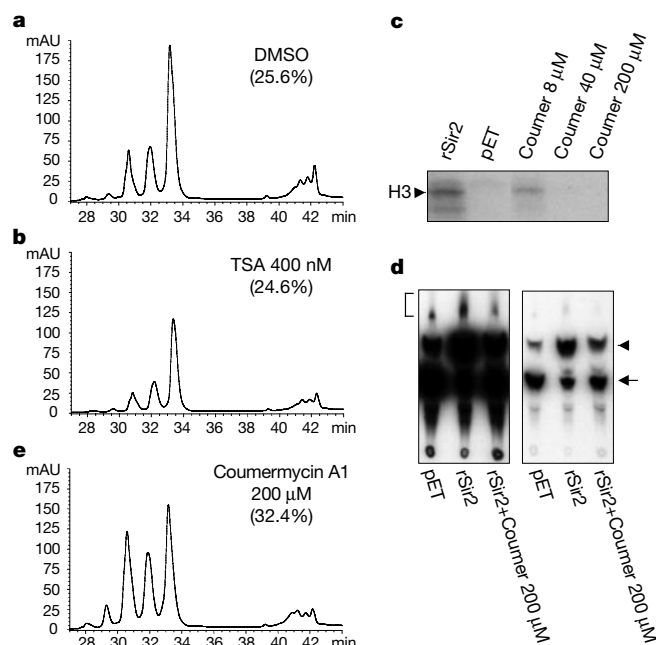


Figure 4 Effects of inhibitors on the deacetylase and ADP-ribosyltransferase activities of recombinant Sir2 (rSir2). HPLC chromatograms of deacetylation reactions at 1 mM NAD in the presence of solvent only (a), 400 nM trichostatin A (TSA) (b), and 200 μ M coumermycin A1 (Coumer) (e) are shown. The calculated efficiencies of the reactions are indicated. The effect of 200 μ M coumermycin A1 on ADP-ribosylation at 1 μ M NAD of the intact histone H3 (c) and the H3 peptide (d) were examined on SDS–PAGE and TLC, respectively. pET corresponds to the vector control. In d, the ADP-ribosylated peptide is indicated by a bracket on a longer exposure (left), and NAD and its hydrolysed product are indicated by an arrow and an arrowhead, respectively, on a shorter exposure (right).

Although NAD and NADH are frequent enzyme co-factors in oxidation/reduction reactions, this is the first example in eukaryotes, to our knowledge, in which NAD drives a distinct enzymatic reaction, in other words, deacetylation by Sir2 of lysines in the N-terminal tails of histones H3 and H4. Our findings are consistent with studies that show the importance of lysines 9 and 14 in the tail of histone H3, and, most critically, of Lys 16 of H4, in silencing⁷⁻⁹. We suggest that deacetylation by Sir2 of Lys 16 of H4, and lysines 9 and 14 of H3 is essential for silencing, and deacetylation of other lysines in these histone tails may be a passive consequence of residence in silent chromatin.

What is the relationship between histone deacetylation and ADP-ribosylation? It has been reported that the ADP-ribosyltransferase activity of Sir2 is essential for silencing *in vivo*¹⁵. Backing this claim was the finding that Sir2 mutant H364Y loses ADP-ribosyltransferase activity *in vitro* and fails to support silencing *in vivo*. However, we find that mutation of this same histidine also greatly reduces the

NAD-dependent deacetylase activity (data not shown), much like our Asn 345 mutant. Moreover, the Gly 270 mutant, which is highly defective for ADP-ribosyltransferase but retains 80% of the wild-type deacetylase activity, is mostly proficient for silencing, recombination suppression and life span extension *in vivo*. Therefore, we surmise that the NAD-dependent deacetylase has a vital role in these Sir2 functions *in vivo*, whereas the ADP-ribosyltransferase does not. Sir proteins are known to move to sites of DNA breaks^{19,20} to aid their repair by non-homologous end joining^{21,22}. We propose that the ADP-ribosyltransferase activity of Sir2 may function in DNA repair. Consistent with this hypothesis are the observations that the ADP-ribosylation of histones occurs when cells are treated with DNA damaging agents²³⁻²⁵ and antibodies against mono-ADP-ribose react with mammalian nuclei only when cells are treated with DNA-damaging agents²⁶.

The NAD requirement of Sir2 for deacetylation suggests that this protein may be a sensor of the energy or oxidation state of cells. It is

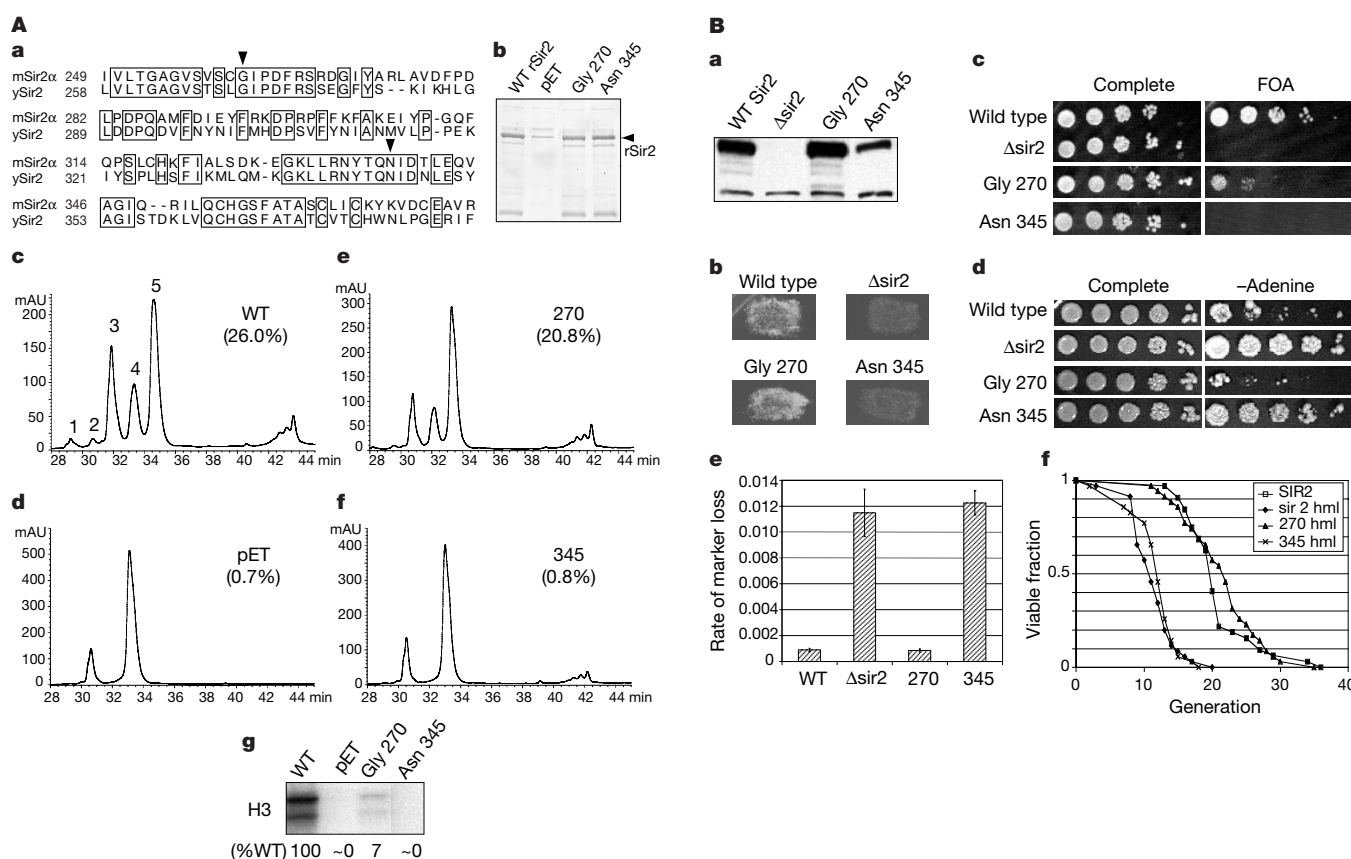


Figure 5 Deacetylation activity of Sir2 is essential for silencing, recombination suppression and life-span extension *in vivo*. **A**, *In vitro* assays of Sir2 enzymatic activities. **a**, Two highly conserved residues (Gly 270 and Asn 345) between yeast Sir2 and mSir2α (indicated by arrowheads) were mutated to alanine. **b**, 6 × His-tagged versions of wild-type and mutant Sir2 proteins along with a vector control (pET) were expressed in *E. coli*, purified over a Ni-NTA column and separated on a 7% polyacrylamide SDS gel stained with Coomassie blue. **c–f**, Effects of mutations in Sir2 on NAD-dependent histone H3 deacetylation activity. Indicated recombinant Sir2 proteins and a vector control were incubated with a peptide corresponding to the di-acetylated N-terminal tail of histone H3 and 1 mM NAD. HPLC chromatograms (absorbance at 220 nm) and efficiencies of the reactions are shown. **g**, Effects of mutations in Sir2 on ADP-ribosylation activity; 1 μg of purified proteins were tested for the ability to ADP-ribosylate intact histone H3 with ³²P-labelled NAD. Lower band corresponds to a degraded histone product. Reaction efficiencies determined by phosphorimager quantitation are shown below. **B**, *In vivo* assays for Sir2 functions. **a**, Western blot of 25 μg yeast whole-cell extracts from wild type, *sir2Δ* and the two mutants was performed using an anti-Sir2 antibody. The upper

band corresponds to Sir2 and a lower background band is included as a loading control. **b**, Silencing at *HMLα* was tested by mating the strains with a tester strain of the opposite mating type and monitoring growth of diploids on selective media. **c**, Telomere silencing was assayed by the ability of strains with telomeric *URA3* to grow on media containing 5-FOA, which is toxic when *URA3* is expressed, but harmless when *URA3* is silenced. Strains were spotted in 10-fold dilutions on media lacking FOA (complete) and containing FOA. **d**, rDNA silencing was monitored in a strain with the *ADE2* marker located within the rDNA array. *RPD3* was disrupted to enhance silencing differences between wild-type and *sir2Δ* strains. *sir2* mutant strains were spotted in 10-fold dilutions on complete and adenine-minus media to monitor effects on silencing. **e**, W303R with *ADE2* at the rDNA was tested for rDNA recombination rates by counting the number of half-sectoring colonies lacking the marker in the first generation after plating. **f**, The mother cell life span of wild-type, *sir2Δ*, and mutants 270 and 345 were determined in *hmlΔ* strains in which *a/α* effects are eliminated. Co-expression of **a** and *α* mating type genes has been shown to shorten life span in W303 (ref. 6).

interesting to note that the anti-ageing regimen of caloric restriction is effective in extending life span in a variety of organisms, including yeast²⁷, *Caenorhabditis elegans*²⁸, rodents²⁹ and probably primates³⁰. We speculate that the slower metabolic rate in energy-deprived cells may exert part of its effect by increasing the availability of NAD, which, in turn, upregulates the deacetylation activity of Sir2 proteins and chromatin silencing. This persistence of genomic silencing may slow ageing-related processes, such as genome instability and inappropriate gene expression. □

Methods

Production of recombinant proteins

The yeast *SIR2* gene or the *mSir2α* full-length cDNA was engineered to be cloned into pET28a vector (Novagen). Site-directed mutations were generated in the plasmid pRS305-SIR2 using the Gene Editor system (Promega) according to manufacturer's instructions. Sequences were verified by Sanger sequencing methods. The mutants were subcloned into pET28a. BL21(DE3) and BL21(DE3)pLysS with an extra copy of arginine transfer RNA gene were transformed with the *SIR2* and *mSir2α* plasmids, respectively. Each transformed bacterial clone was induced in 1 mM IPTG at 37 °C for 1 h. The induced 6 × His-tagged proteins were purified with Ni-NTA agarose under native conditions. The control eluate was prepared from a bacterial clone carrying pET28a vector only. The recombinant proteins were aliquoted and kept at -70 °C.

Deacetylation and ADP-ribosylation assays

The typical reaction of Sir2 deacetylase activity was performed in 50 μl of buffer containing 50 mM Tris-HCl, pH 9.0, 4 mM MgCl₂, 0.2 mM dithiothreitol (DTT), a variable concentration of unlabelled nicotinamide adenine dinucleotide (NAD) or NAD derivatives (Sigma, 5–10 μg of the purified recombinant Sir2 proteins and 10 μg of the N-terminal tail peptide of histone H3 (residues 1–20 + Cys; ARTKQTAR(AcK)STG-G(AcK)APRKQLC) di-acetylated at positions 9 and 14 or of *Tetrahymena* histone H4 (residues 2–19 + Cys; AGG(AcK)GG(AcK)GMG(AcK)VGA(AcK)RHSC) tetra-acetylated at positions 5, 8, 12 and 16 (Upstate Biotechnology). The starting peptide material of H3 contains a contaminant with a relative molecular mass 100 smaller, which also showed exactly the same patterns of deacetylation (data not shown). To detect the ADP-ribosylation activity, 8 μCi of NAD 5'-[α-³²P]triphosphate (~1,000 Ci mmol⁻¹, Amersham Pharmacia) was added to the same reaction containing 1 μM unlabelled NAD. Four micrograms of intact histone H3 protein (Roche Molecular Biochemicals) was also used for this assay. All reaction mixtures were incubated at room temperature for 1 h. Trichostatin A and coumermycin A1 (Sigma) were prepared in dimethylsulphoxide (DMSO), and 5 μl of solvent or inhibitor was added to the reactions before adding Sir2 proteins.

Analysis of deacetylated or ADP-ribosylated products

After the incubation, the products were precipitated at -20 °C overnight by adding 50 μl of distilled water and 25 μl of 100% trichloroacetic acid (TCA) solution. For HPLC, the precipitates were reconstituted in 5% CH₃CN and 0.1% trifluoroacetic acid (TFA) and run through a gradient concentration of 0.05% TFA to 0.043% TFA plus 80% CH₃CN on Hewlett Packard Model 1100 HPLC system with 214TP52 column (Vydac). The chromatograms (absorbance measured at 210 nm) were digitally recorded and analysed by Hewlett Packard ChemStation system (version A.06.03(509)). Fractions of samples were collected every 1 min by Gilson Fraction Collector Model 203. Peptide sequencing was done by the Applied Biosystems Precise 494 HT protein sequencing system. PTH amino-acid chromatograms were recorded and analysed by the ABI Model 610A2.1 data integration/analysis system. Electron-spray mass spectroscopy was done on the PE Sciex Model API365 system. The mass spectroscopy data were analysed by the BioMultiView program (version 1.3.1). To detect the ADP-ribosylated intact H3 protein, the pellets were dissolved in 15 μl of Laemmli's sample buffer, boiled for 1.5 min and electrophoresed in a 10–20% gradient SDS-PAGE gel. The gel was stained with Coomassie Brilliant Blue to check equal loading of histones, dried and exposed to Kodak X-OMAT film. To analyse the ADP-ribosylated H3 peptides, 10 μl of each reaction mixture was spotted on a cellulose TLC plate (EM Science). Chromatography was carried out for 9–10 h in a TLC chamber containing a 65:53:2:29 mixture of isobutyric acid, pyridine, acetic acid, butanol and water. The plate was dried and exposed to Kodak X-OMAT film. Peptide spots were checked by ninhydrin staining.

Molecular cloning of mSir2α

The mouse 15-day embryo 5'-STRETCH PLUS cDNA library (Clontech) was screened with the expressed sequence tag cDNA fragment of AA199012 as a probe. Five positive clones were obtained from roughly one-million independent plaques. One of the five clones contained a 3.9-kb cDNA fragment. The nucleotide sequence of this fragment was determined with an Applied Biosystems 374 automated sequencer. We concluded that the cDNA clone that we isolated encodes the full-length mSir2α protein, because the *in vitro* translated protein from this cDNA clone had an *M_r* (110–120K) indistinguishable from that of the protein in the mouse NIH3T3 extract recognized by a specific polyclonal antibody against a N-terminal portion of mSir2α (data not shown). The amino-acid sequences of mSir2α and other *SIR2* family members were aligned in the Clustal X program and a phylogenetic tree was generated by using the NJPLOT program (see Supplementary Information).

Strains, plasmids and antibodies

All strains used were derivatives of W303a *sir2Δ*: W303R *sir2Δ*(*MATa*, *ade2-1*, *leu2-3,112*, *trp1-1*, *ura3-52*, *his3-11*, *sir2::TRP1*, *rDNA-ADE2*), W303RT *sir2Δ*(*MATa*, *ade2-1*, *leu2-3,112*, *trp1-1*, *ura3-52*, *his3-11*, *rad5-535*, *sir2::TRP1*, *rDNA-ADE2*, *TELVIII-URA3*), and W303R *sir2Δ/rpd3Δ*(*MATa*, *ade2-1*, *leu2-3,112*, *trp1-1*, *ura3-52*, *his3-11*, *rdp3::URA3*, *sir2::TRP1*, *rDNA-ADE2*). We used pRS305-SIR2, an integrating plasmid that contains *SIR2* driven by its native promoter. Mutant *SIR2* genes were also cloned into these vectors. *SIR2* and mutant *sir2* strains were generated by cutting the plasmid within the *LEU2* gene and integrated using standard yeast transformation protocols. Rabbit antibodies against Sir2 were generated by using full-length recombinant Sir2 purified under denaturing conditions.

Silencing, life span and rDNA recombination assays

To test silencing at the telomeres and rDNA, 10-fold dilutions of the derivatives of W303RT and W303R *Drp3* were spotted on media containing 5-FOA and media lacking adenine, respectively. To assay for HM silencing, W303R derivatives were spotted onto YPD with the tester strain CKy20 (*MATα*, *arg1*, *tsm11*) and after overnight growth were replica plated to minimal media with no supplemented amino acids. Life span and rDNA recombination rates were measured as described⁶.

Received 2 December 1999; accepted 26 January 2000.

- Rine, J. & Herskowitz, I. Four genes responsible for a position effect on expression from *HML* and *HMR* in *Saccharomyces cerevisiae*. *Genetics* **116**, 9–22 (1987).
- Gottschling, D. E., Aparicio, O. M., Billington, B. L. & Zakian, V. A. Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* **63**, 751–762 (1990).
- Bryk, M. *et al.* Transcriptional silencing of Ty1 elements in the RDN1 locus of yeast. *Genes Dev.* **11**, 255–269 (1997).
- Smith, J. S. & Boeke, J. D. An unusual form of transcriptional silencing in yeast ribosomal DNA. *Genes Dev.* **11**, 241–254 (1997).
- Gottlieb, S. & Esposito, R. E. A new role for a yeast transcriptional silencer gene, *SIR2*, in regulation of recombination in ribosomal DNA. *Cell* **56**, 771–776 (1989).
- Kaerberlein, M., McVey, M. & Guarente, L. The *SIR2/3/4* complex and *SIR2* alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Genes Dev.* **13**, 2570–2580 (1999).
- Thompson, J. S., Ling, X. & Grunstein, M. Histone H3 amino terminus is required for telomeric and silent mating locus repression in yeast. *Nature* **369**, 245–247 (1994).
- Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M. & Grunstein, M. Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: A molecular model for the formation of heterochromatin in yeast. *Cell* **80**, 583–592 (1995).
- Braunstein, M., Sobel, R. E., Allis, C. D., Turner, B. M. & Broach, J. R. Efficient transcriptional silencing in *Saccharomyces cerevisiae* requires a heterochromatin histone acetylation pattern. *Mol. Cell. Biol.* **16**, 4349–4356 (1996).
- Braunstein, M., Rose, A. B., Holmes, S. G., Allis, C. D. & Broach, J. R. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes Dev.* **7**, 592–604 (1993).
- Sinclair, D. A. & Guarente, L. Extrachromosomal rDNA circles—a cause of aging in yeast. *Cell* **91**, 1–20 (1997).
- Brachmann, C. B. *et al.* The *SIR2* gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability. *Genes Dev.* **9**, 2888–2902 (1995).
- Tsang, A. W. & Escalante-Semerena, J. C. CobB, a new member of the SIR2 family of eucaryotic regulatory proteins, is required to compensate for the lack of nicotinate mononucleotide:5,6-dimethylbenzimidazole phosphoribosyltransferase activity in *cobT* mutants during cobalamin biosynthesis in *Salmonella typhimurium* LT2. *J. Biol. Chem.* **273**, 31788–31794 (1998).
- Frye, R. A. Characterization of five human cDNAs with homology to yeast SIR2 gene: Sir2-like proteins (Sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity. *Biochem. Biophys. Res. Commun.* **260**, 273–279 (1999).
- Tanny, J. C., Dowd, G. J., Huang, J., Hilz, H. & Moazed, D. An enzymatic activity in the yeast Sir2 protein that is essential for gene silencing. *Cell* **99**, 735–745 (1999).
- Yoshida, M., Kijima, M. & Akita, M. & Beppu, T. Potent and specific inhibition of mammalian histone deacetylase both *in vivo* and *in vitro* by Trichostatin A. *J. Biol. Chem.* **265**, 17174–17179 (1990).
- Taunton, J., Hassig, C. A. & Schreiber, S. L. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**, 408–411 (1996).
- Banasik, M. & Ueda, K. Inhibitors and activators of ADP-ribosylation reactions. *Mol. Cell. Biochem.* **138**, 185–197 (1994).
- Mills, K. D., Sinclair, D. A. & Guarente, L. *MEC1*-dependent redistribution of the Sir3 silencing protein from telomeres to DNA double-strand breaks. *Cell* **97**, 609–620 (1999).
- Martin, S. G., Laroche, T., Suka, N., Grunstein, M. & Gasser, S. M. Relocalization of telomeric Ku and SIR proteins in response to DNA strand breaks in yeast. *Cell* **97**, 621–633 (1999).
- Boulton, S. J. & Jackson, S. P. Identification of a *Saccharomyces cerevisiae* Ku80 homolog: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Res.* **24**, 4639–4648 (1998).
- Tsukamoto, Y., Kato, J. & Ikeda, H. Silencing factors participate in DNA repair and recombination in *Saccharomyces cerevisiae*. *Nature* **388**, 900–903 (1997).
- Adamietz, P. & Rudolph, A. ADP-ribosylation of nuclear proteins *in vivo*: identification of histone H2B as a major acceptor for mono- and poly(ADP-ribose) in dimethyl sulfate-treated hepatoma AH7974 cells. *J. Biol. Chem.* **259**, 6841–6846 (1984).
- Kreimeyer, A., Wielckens, K., Adamietz, P. & Hilz, H. DNA repair-associated ADP-ribosylation *in vivo*: modification of histone H1 differs from that of the principal acceptor proteins. *J. Biol. Chem.* **259**, 890–896 (1984).
- Pero, R. W., Holmgren, K. & Persson, L. Gamma-radiation induced ADP-ribosyltransferase activity and mammalian longevity. *Mutat. Res.* **142**, 69–73 (1985).
- Meyer, T. & Hilz, H. Production of anti-(ADP-ribose) antibodies with the aid of a dinucleotide-pyrophosphatase-resident hapten and their application for the detection of mono(ADP-ribose)lated polypeptides. *Eur. J. Biochem.* **155**, 157–165 (1986).

27. Muller, I., Zimmermann, M., Becker, D. & Flomer, M. Calendar life span versus budding life span of *Saccharomyces cerevisiae*. *Mech. Ageing Dev.* **12**, 47–52 (1980).
28. Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95**, 13091–13096 (1998).
29. Weindruch, R. H., Walford, R. L., Fligiel, S. & Guthrie, D. The retardation of aging in mice by dietary restriction: Longevity, cancer, immunity, and lifetime energy intake. *J. Nutr.* **116**, 641–654 (1986).
30. Roth, G. S. Caloric restriction in primates: will it work and how will we know? *J. Am. Geriatr. Soc.* **47**, 896–903 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank R. Cook, A. Park, and H. Amoroso at the MIT Biopolymers lab for the HPLC and electron-spray mass spectroscopy, and P. Matsudaira for the MALDI mass spectroscopy. We also thank S. Inamoto for bacterial strains, and H. Tissenbaum and E. Ford for comments on the paper. This work was funded by The Human Frontier Science Program Organization Long-Term Fellowship to S. L., a NIH predoctoral grant to C.A. and M.K., and grants from the NIH, Seaver Foundation, Ellison Medical Foundation, and Howard and Linda Stern Fund to L.G.

Correspondence and requests for materials should be addressed to L.G. (e-mail: leng@mit.edu).

The structures of HslU and the ATP-dependent protease HslU–HslV

Matthias Bochtler^{*†}, Claudia Hartmann^{*†}, Hyun Kyu Song^{*}, Gleb P. Bourenkov[‡], Hans D. Bartunik[‡] & Robert Huber^{*}

^{*} Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, Planegg, Germany

[‡] Arbeitsgruppen für strukturelle Molekularbiologie der Max-Planck-Gesellschaft, Notkestr. 85, D-22603 Hamburg, Germany

[†] These authors contributed equally to this work

The degradation of cytoplasmic proteins is an ATP-dependent process¹. Substrates are targeted to a single soluble protease, the 26S proteasome^{2,3}, in eukaryotes and to a number of unrelated proteases in prokaryotes^{4,5}. A surprising link emerged with the discovery of the ATP-dependent protease HslVU (heat shock locus VU)^{6–8} in *Escherichia coli*. Its protease component HslV shares ~20% sequence similarity⁶ and a conserved fold⁹ with 20S proteasome β -subunits. HslU is a member of the Hsp100 (Clp) family of ATPases. Here we report the crystal structures of free HslU and an 820,000 relative molecular mass complex of HslU and HslV—the first structure of a complete set of components of an ATP-dependent protease. HslV and HslU display sixfold symmetry, ruling out mechanisms of protease activation that require a symmetry mismatch between the two components. Instead, there is conformational flexibility and domain motion in HslU and a localized order–disorder transition in HslV. Individual subunits

of HslU contain two globular domains in relative orientations that correlate with nucleotide bound and unbound states. They are surprisingly similar to their counterparts in *N*-ethylmaleimide-sensitive fusion protein^{10,11}, the prototype of an AAA-ATPase. A third, mostly α -helical domain in HslU mediates the contact with HslV and may be the structural equivalent of the amino-terminal domains in proteasomal AAA-ATPases.

We expressed and purified histidine-tagged variants of HslV and HslU from *E. coli*. Mixing the two components was sufficient to recover ATP-dependent protease activity, as assessed by enzymatic assay with both the chromogenic peptide Z-GGL-AMC⁷ and radio-labelled ¹⁴C-methylcasein¹²; the latter was used as a model system for large, denatured proteins (data not shown). Three crystal forms grew (Fig. 1; and Table 1). The crystals of space group *P*₆₃22 contain free HslV and HslV–HslU fully bound with AMP-PNP (5'-adenylylimidetriphosphate) on 32-point symmetry positions (Fig. 1a). In the absence of both HslV and exogenous nucleotide, HslU crystallized in space group *P*₂₁2₁2₁, with a full hexamer in the asymmetric unit. Four subunits are complexed with nucleotide, presumably from the cellular environment. The remaining two subunits are located on opposite sides of the ring and contain bound sulphate from the crystallization buffer (Fig. 1b). The addition of the non-hydrolysable AMP-PNP induced the growth of trigonal crystals under identical conditions with two independent hexamers located on threefold crystallographic axes with nucleotide bound to alternating subunits on the ring (Fig. 1c). Empty subunits have been modelled without sulphate and show local main-chain conformational changes of the phosphate-binding loop (P-loop).

High sequence similarity predicts that there is a conserved fold in all members of the Hsp100 (Clp) family of ATPases. In contrast, the identification of the Hsp100s as AAA proteins¹³ has been controversial¹⁴. Comparison of the HslU structure with the only atomic-resolution structure of an AAA protein available so far^{10,11} strongly supports this classification (see Fig. 2). HslU is folded into three distinct domains. We observe the same topology in the N-terminal domain (D2) of *N*-ethylmaleimide-sensitive fusion (NSF) protein and the N-terminal domain of HslU. In HslU, this domain comprises residues S2–K109 and I244–L332 and will be referred to as the N domain. Residues M110 to A243 are missing in other Hsp100 proteins and emerge from the globular domain. The structural equivalence of the HslU and NSF D2 backbones is greatest (r.m.s. deviation of C α bond lengths = 0.7 Å) for a central, parallel β -sheet that forms the core of the protein in both structures, having strand order β N2, β N3, β N6, β N1, β N7 in HslU (see Fig. 2; and Table 2). C α carbons of corresponding residues in the sequence alignment are nearly coincident, with the exception of residues PFIK of strand β N2 in HslU, which correspond to residues FIKI in NSF. Lysine 80 is extended towards the nucleotide in HslU. The central β -sheet in both HslU and NSF is sandwiched in between two layers of α -helices as in other P-loop proteins. The P-loop is located between strand β N1 and helix α N4 in HslU and is in contact with nucleotide as expected^{15,16}. As in NSF, and unlike in other P-loop triphosphate

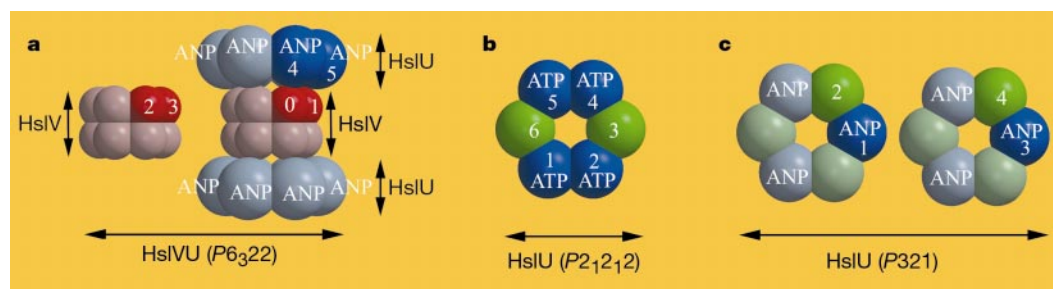


Figure 1 Summary of the three crystal forms (a–c) that were used for structure determination. Subunits in the respective asymmetric units are numbered 1–6.

The intermediate domain (I domain) emerges from the N domain with residue M110 and returns to it with residue A243. It is ordered and defined by electron density to varying extents in the subunits, as shown in Table 2. Where the intermediate domain is ordered, it is a loosely folded structure with coiled helices, one from residues M110–I136 (α I1/ α I2), and the other from residue K217–

b. Stereo diagram of NSF D2.

‡ The backtransformation *B*-factor was calculated after ten cycles of twofold averaging.

The carboxy-terminal domain (C domain, residues 333–443) is a four-helix bundle (α C1/ α C2, α C3, α C4, α C5/ α C6) with a large bulge at residues T345–S350 and a double-stranded parallel β -sheet formed by β -strands β C1 and β C2. This forms a planar surface and is a prominent crystal contact site in all three crystal forms. The helix topology resembles that of the C domain in NSF, which is much smaller as it lacks the accessory structures mentioned above. The FIL (441–443) C terminus of HslU, which is conserved among the HslU family, is buried inside the domain with its charge balanced by R394 and by R329 of the adjacent subunit.

In our HslV–HslU crystals, HslV is in contact with HslU along and perpendicular to its sixfold axis. A complex was assigned from considerations of stoichiometry and symmetry, such that HslV and HslU hexamers form a *D*₆ symmetric assembly. It is a doubly capped species, as in electron microscopy images¹⁸. Our data confirm earlier data on the diameter of HslU, but disagree with the electron microscopy data on the overall size of the HslV–HslU complex. The crystal structure places the bulk of the scattering mass of HslU 5 nm away from the protease, separated by the I domains of HslU (see Fig. 3a). As a result, the overall dimensions of the complex are 25 nm in length and 12 nm in diameter. The compact shape of the molecule in electron microscopy images may be attributed to a collapsed species, but could also indicate an alternative mode of association that has the N and C domains of HslU in contact with HslV. The electron microscopy data show the presence of both hexameric and heptameric HslU species, indicating alternative modes of assembly also on this level¹⁸. The role of the I domains as spacers is maintained in both HslU crystal forms, where they span roughly the same distance between neighbouring HslU rings.

In our crystals of HslV–HslU, residues 140–149 of the I domain of HslU contact residues around 39 and 66 in HslV (see Fig. 3b). The interactions are not precisely defined, as the contacting HslU segment and the HslV segment at residue 66 are partly disordered

in the complex. Neither the sequences of the contacting segment of HslV nor those of HslU give a clue to the specificity of this interaction. We note that in both ours and other hands⁷, the complex in solution is rather unstable and does not survive most chromatographic procedures.

A direct comparison of the bound and free HslV species in the HslV–HslU crystals suffers from ill-defined electron density for the free species. A comparison of the HslV component of the complex with previously analysed isolated HslV shows small shifts of subunits within the dodecamer probably as a consequence of different crystal packing constraints and, prominently, a disordering of segments 65–69 and 84–93 (see Fig. 3b). The latter segment is an α -helix lining the entrance channel into the HslV protease chamber.

The nucleotide in HslU subunits is in contact with residues from both the N and the C domains (see Fig. 4). Residues V61, I17 and I18 from the N domain and residues L335, I343 and A392 from the C domain form a hydrophobic pocket that accommodates the purin ring. The main-chain residues H16–I18 and G60–V61 supply hydrogen acceptors for the adenine base. The sugar is sandwiched in between E56 and H396. A number of residues of the N domain contact the phosphate moiety of ATP. The peptide amido groups of G60–E65 are close enough to the α - and β -phosphates to form hydrogen bonds. In addition, the phosphates are coordinated by a set of polar side chains. These include K63 of the Walker A motif, K80, which lies within hydrogen-bonding distance of the γ phosphate, and D256. Threonine 59 and T64 face the γ -phosphate from opposing sides. The C domain interacts with the β - and γ -phosphates through R393, a residue that is highly conserved among AAA-ATPases¹³ and that is part of the signature sequence V of Clp ATPases¹⁹. Two acidic residues of the adjacent subunit, E286 and E321, are also in contact with nucleotide. Although acidification of the phosphate-binding pocket as a result of oligomer assembly could be the trigger for ATP-hydrolysis in HslU, this

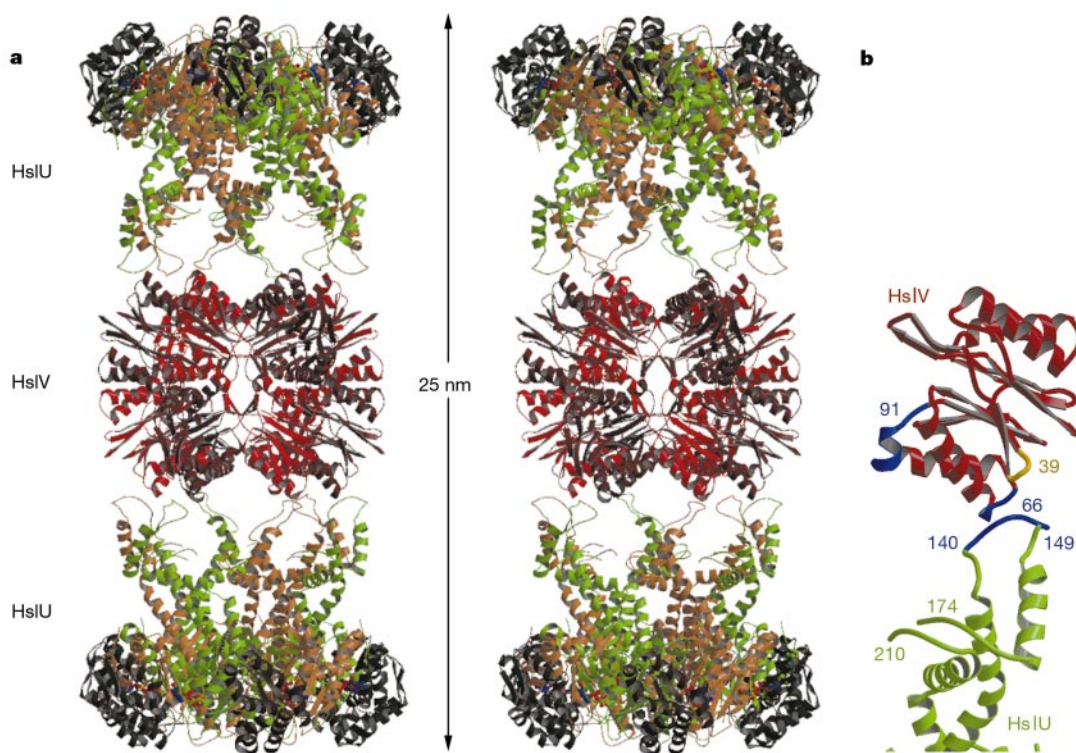


Figure 3 HslV–HslU crystal complex. **a**, Ribbon plot (MOLSCRIPT) representation of the HslV–HslU crystal complex. Data from the HslU crystals has been included to model intermediate domains that were disordered in the HslV–HslU crystals. **b**, Close up view showing details of the contact region between HslV and HslU, with the HslV subunit taken

from the lower right corner of the HslV particle. Residues of HslV and HslU that are disordered in the HslV–HslU complex are shown in blue. An additional site of contact in HslV that remains ordered in the HslV–HslU complex is shown in yellow.

conclusion has to be treated with caution. Neither of the acidic residues is conserved in other AAA proteins, and NSF D2 has a lysine (K639) precisely in the position of E321. Arginine 325, a highly conserved residue¹³ and the homologue of a putative arginine finger²⁰ in FtsH, is substantially further away (~8 Å) and may act indirectly.

Negative cooperativity of ATP binding in adjacent subunits is suggested by the orthorhombic (see Fig. 1b) and trigonal (see Fig. 1c) HslU crystals. It might explain the widely different apparent Michaelis constant (K_m) for peptide and protein hydrolysis¹². A decrease in ATPase activity with decreasing ATP saturation could be deduced from a superposition of the nucleotide-bound subunits. We find a retraction of the mentioned acidic groups from the γ -phosphate when nucleotide is absent in an adjacent subunit (data not shown). Unfortunately, electron density for the γ -phosphate itself is ill-defined in our structures even for the non-hydrolysable AMP-PNP analogue, owing to the absence of Mg^{2+} which is caused by an excess of EDTA in the crystallization conditions for the HslV–HslU (Fig. 1a) and orthorhombic HslU (Fig. 1b) crystals.

The absence of nucleotide leads to a relaxation of the structure, with the N and C domains rotating away from each other as rigid entities (see Fig. 2a). Looking down the sixfold axis of HslU towards the HslV complex, the C domain of one subunit is in contact with the N domain of the neighbouring subunit in counterclockwise orientation. Unexpectedly, this contact is so tight that the two domains, which are not covalently linked, behave as one rigid unit even in the context of the HslU oligomer. It appears that the

movement of the N domain is entirely driven by the change in orientation relative to the adjacent C domain (see Fig. 5). The acidic residues E286 and E321 of the N domain, which were directly in contact with nucleotide before its hydrolysis and dissociation, are so close to the centre of the overall rotation that they hardly move. In contrast, strands β N4 and β N5 are displaced by nearly 10 Å. The domain movements shift the I domain and may serve to dissociate sticky substrates and deliver them to HslV. In addition, they lead to a loss of proper sixfold symmetry in partially nucleotide-bound HslU particles. The resulting mismatch between the distorted sixfold symmetry of HslU and the proper sixfold symmetry of HslV should reduce the affinity of the two particles. Consistent with this hypothesis, we observe HslU with AMP-PNP bound to all subunits in the HslV–HslU co-crystals (see Fig. 1a) and partially AMP-PNP-bound HslU molecules in both HslU crystal forms (see Fig. 1b, c).

The crystal structures constrain models for substrate recognition and translocation. It is incompatible with the suggested presence of PDZ-domains²¹ in the C-terminal part of the molecule. The previously characterized “sensor and substrate discrimination” (SSD) domain²² coincides with the C domain of the crystal structure. Unlike in NSF D2, this domain is substantially involved in inter-subunit contacts and does not protrude from the ring. Its molecular surfaces face either adjacent subunits or solvent. There is no area of contact with the HslU cavity. We imagine a major role for the I domains in substrate recognition. The I domains, which are mostly disordered in three different crystal forms except when they make contacts with crystallographic neighbours (see Tab. 2b), emerge

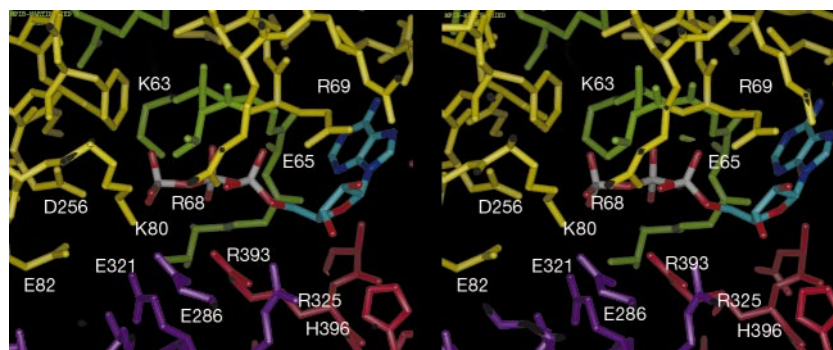


Figure 4 Representation of the nucleotide-binding site in HslU. The nucleotide is bound at the interface between two subunits. The N and C domains of the subunit that contributes most interactions are shown in yellow and red, respectively. Although the P-loop is part of

the N-domain of this subunit, it has been highlighted in green. Residues of the adjacent subunit are shown in purple.

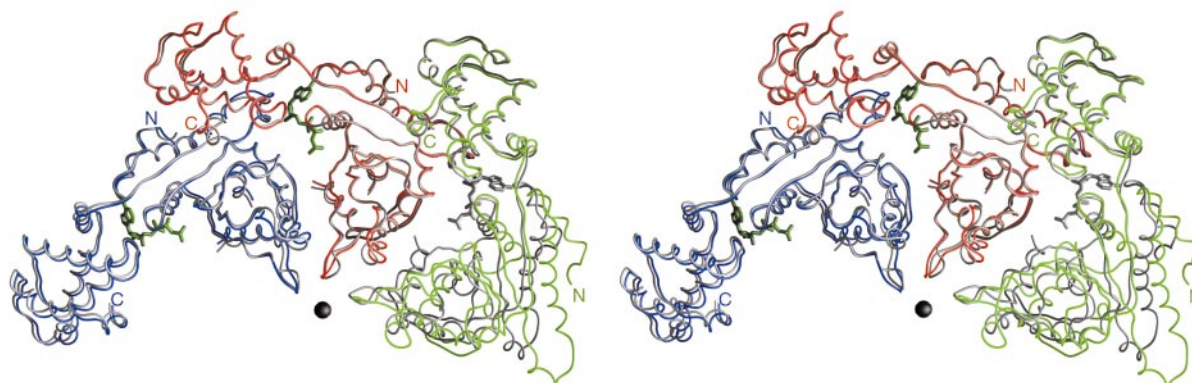


Figure 5 View of HslU along the sixfold axis (indicated by a circle). Three subunits on the ring and all I domains have been omitted for clarity. Fully ligand-bound HslU (from the P6322 crystals) shown in grey is superimposed on partially ligand-bound HslU (from the

P21212 crystals) shown in blue, red and green to distinguish individual subunits. Green nucleotide is present in both species, grey nucleotide only in the fully ligand-bound form. Note the movement of the N domain on the left side of the drawing.

Table 2 HslU secondary structure assignment and list of disordered residues in HslU in the three crystal forms

HslU secondary structure assignment*
α N1(P6–K15), α N2(D21–Q39), α N3(E42–E47), β N1(I53–I56) [V547–E550], α N4(K63–A74), β N2(F78–V81) [I573–C576], α N5(A83–T87), α N6(V96–K109), α 1(M110–K118), α 2(Y121–I136), α 3(T146–G166), β 1(L168–E174), β 2(K210–L216), α 4(K217–L233), α 5(P236–A243), α N7(I244–H250), β N3(I252–I255) [C607–V610], α N8(S268–E286), β N4(T289–T292), β N5(G295–K298), β N6(L303–G308) [L649–T654], α N9(P320–G324), β N7(I328–E331) [T671–H674], α C1(T337–E346), α C2(I351–E362), β C1(N365–F368), α C3(D370–S386), α C4(A392–L413), β C2(N417–N420), α C5(A422–A434), α C6(E436–I442)
Disordered residues in HslU in the three crystal forms†
HslVU (P ₆ ,22)
140–152, 167–215 (4); 114–157, 164–233 (5)
HslU (P ₂ ,2,2)
128–233 (1); 139–149, 177–212 (2); 118–230 (3); 129–229 (4); 142–150, 177–212 (5); 118–221 (6)
HslU (P321)
175–209 (1); 135–222 (2); 175–209 (3); 135–221 (4)

* Structurally equivalent residues in NSF are indicated in square brackets for the central, parallel β -sheet.

† Chain identifiers are given in parentheses (see Fig. 1).

teeth-like from a base plate made up of the N and C domains and may assist entrapment of substrate molecules of variable size and shape, similar to the action of a grab dredge. Their entirely disordered segment I175–Q209 may act as a bait and/or an adaptor to substrate molecules and promote binding. Substrates inside HslU may be destabilized and partially unfolded, as has been observed in ClpA²³. Association of HslU with HslV leads to the observed local disordering of the HslV entrance channel (see Fig. 3b). We note that in our crystallization conditions HslVU is expected to be less active than in the presence of ATP and low salt¹². Therefore, the role of the described structures in the reaction cycle requires further characterization.

Atomic-resolution structures have been reported for 20S proteasomes from archaeae²⁴ and yeast²⁵ and for the mammalian, ATP-independent proteasome activator PA28 (ref. 26). In contrast, only biochemical and electron microscopy data are available for the 19S regulatory cap. These data show that the particle can be dissociated into subcomplexes called lid and base. Along with other subunits, the base contains the six proteasomal AAA-ATPases²⁷. They can be aligned with the N and C domains of HslU, as expected from our structure-based classification of HslU as an AAA protein. Unexpectedly, there may even be a counterpart for the I domain of HslU: it shares 14% identity and 37% homology with residues 11–141 of the N terminus of MSS1. We suggest that the proteasomal ATPases are structurally similar to HslU but have the I domain fused to the N terminus, which is nearby. We expect a ring-like, oligomeric structure, as in HslU, with the non-ATPase subunits present to modulate its activity. Our structure does not make any predictions for the lid complex which must have evolved with the ubiquitin system. □

Methods

Protein purification

HslV was purified as previously described. HslU from *E. coli* was cloned into pET12b with an N-terminal MHHHHHH tag and expressed under the transcriptional control of the T7 promoter in BL21(DE3). Cells were collected after a 4-h induction with IPTG, resuspended in 10 mM Tris-HCl pH 7.5 and lysed by sonification in the presence of 1 mM phenylmethylsulphonyl fluoride (PMSF) and of 20 mg lysozyme and 1 mg DNaseA per litre of culture. The lysate was clarified by ultracentrifugation and applied to a chelating Sepharose column (ϕ = 2.5 cm, h = 5 cm, Pharmacia) loaded with 2 mg ml⁻¹ of Zn²⁺. HslU eluted between 20 and 50 mM imidazole in buffer A (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.002% NaN₃). HslU-containing fractions were pooled, supplemented with 5 mM ethylenediaminetetraacetic acid (EDTA) and applied to a Q-sepharose FF column (ϕ = 3.5 cm, h = 7 cm, Pharmacia) equilibrated with 200 mM NaCl in buffer B (20 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM NaN₃). HslU eluted at 300 mM NaCl. Fractions were pooled, concentrated and subjected to a final gel-filtration step in buffer B on a Sephacryl S300 column.

Crystallization

Crystals of HslU grew at room temperature within a few days by sitting-drop vapour diffusion against a reservoir containing 100 mM sodium cacodylate pH 6.5, 15% glycerol, 10.5% polyethylene glycol (PEG) 8K and 500 mM (NH₄)₂SO₄. Drops initially contained 2 μ l reservoir solution, 2 μ l protein at 16 mg ml⁻¹ in buffer B. The addition of 0.5 μ l of 1.0 M guanidinium chloride and of 1 μ l detergent octaethylene glycol monododecyl ether (C₁₂E₈) improved the quality of the crystals. Without the addition of nucleotide, we could grow orthorhombic crystals that diffracted to 3 Å resolution. The addition of 1 mM AMP-PNP and of 5 mM MgCl₂ induced the growth of trigonal crystals under otherwise identical conditions. Both crystal forms could be frozen in reservoir buffer mixed with 87% glycerol in the ratio 17:3. For crystallization of the protein complex, a 20 mg ml⁻¹ solution of HslU in buffer B with 1 mM AMP-PNP was prepared. It was mixed in a 2:1 ratio with a 16 mg ml⁻¹ solution of histidine-tagged HslV in 300 mM NaCl, 20 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM NaN₃. Crystals were grown at room temperature in sitting drops, equilibrating 2 μ l of protein solution and reservoir buffer each against the reservoir buffer that contained 100 mM MES pH 6.3 and 2.0 M sodium acetate. A 17:3 mixture of reservoir solution and D(-)-2,3-butanediol served as cryoprotectant.

Structure determination

All diffraction data were recorded at BW6, DESY, Hamburg (see Table 1). The HslV–HslU structure was solved by a combination of the molecular replacement and MAD phasing methods. Anomalous data were collected with crystals soaked with either tantalum bromide (TaBr₅) or thiomersal (C₆H₅HgNaO₂S). Both derivatives were too weak to be localized by conventional Patterson search methods. The clue to structure determination was our assumption that one HslV particle occupied a position of 32-point symmetry in the HslV–HslU crystals. This assumption was supported by the self-rotation function and the strong pseudo-origin peaks at (2/3,1/3,1/2) and (1/3,2/3,1/2). It could be confirmed by molecular replacement methods (AMORE²⁸) with the HslV dodecamer as the search model. Although the two independent subunits of HslV represented less than a quarter of the total scattering mass in the asymmetric unit, their placement allowed the calculation of partial phases that were good enough to identify candidate heavy atom positions in both anomalous and dispersive difference Fourier maps. All sites that could be confirmed by cross-vector verification (RSPS) were refined and used for initial phasing (MLPHARE²⁸). The figure of merit was 0.53 to 2.8 Å resolution and could be improved to 0.55 by phase combination with partial model phases. The corresponding electron density map allowed the tracing of both HslU subunits with the exception of the intermediate domain, and showed the presence of a free HslV dodecamer in the crystals. Phases were improved by solvent flattening and cyclic averaging of the HslV–HslU electron density. The HslU model from the HslV–HslU crystals without the intermediate domains could be positioned in both HslU crystal forms by molecular replacement (AMORE). The originally weak signal improved markedly after allowing for mobility in the hinge region around residue 333. Cyclic averaging was used to improve phases in all three crystal forms. Models were built with FRODO²⁹ and refined with X-PLOR (P321) or CNS³⁰ (P₆,22 and P₂,2,2) with Engh & Huber parameters.

Received 21 September; accepted 23 December 1999.

- Goldberg, A. L. & St. John, A. C. Intracellular protein degradation in mammalian and bacterial cells: part 2. *Annu. Rev. Biochem.* **45**, 747–803 (1976).
- Hershko, A. & Ciechanover, A. The ubiquitin system for protein degradation. *Annu. Rev. Biochem.* **61**, 761–807 (1992).
- Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
- Maurizi, M. R. Proteases and protein degradation in *Escherichia coli*. *Experientia* **48**, 178–201 (1992).
- Gottesman, S., Wickner, S. & Maurizi, M. R. Protein quality control: triage by chaperones and proteases. *Genes Dev.* **11**, 815–823 (1997).
- Chuang, S. E., Burland, V., Plunkett, G., Daniels, D. L. & Blattner, F. R. Sequence analysis of four new heat shock genes constituting the hslu and hslv operons in *Escherichia coli*. *Gene* **134**, 1–6 (1993).
- Rohrwild, M. *et al.* HslV-HslU: a novel ATP-dependent protease complex in *Escherichia coli* related to the eukaryotic proteasome. *Proc. Natl Acad. Sci. USA* **93**, 5808–5813 (1996).
- Missiakis, D., Schwager, E., Betton, J. -M., Georgopoulos, C. & Raina, S. Identification and characterization of HslV HslU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in *Escherichia coli*. *EMBO J.* **15**, 6899–6909 (1996).
- Bochtler, M., Ditzel, L., Groll, M. & Huber, R. Crystal structure of heat shock locus V (HslV) from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **94**, 6070–6074 (1997).
- Lenzen, C. U., Steinmann, D., Whiteheart, S. W. & Weiss, W. I. Crystal structure of the hexamerization domain of N-ethylmaleimide-sensitive fusion protein. *Cell* **94**, 525–536 (1998).
- Yu, R. C., Hanson, P. I., Jahn, R. & Brünger, A. T. Structure of the ATP-dependent oligomerization domain of the N-ethylmaleimide sensitive factor complexed with ATP. *Nature Struct. Biol.* **5**, 803–811 (1998).
- Huang, H. -C. & Goldberg, A. L. Proteolytic activity of the ATP-dependent protease HslVU can be uncoupled from ATP-hydrolysis. *J. Biol. Chem.* **272**, 21364–21372 (1997).
- Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: a class of chaperone-like ATPases associated with the assembly, operation and disassembly of protein complexes. *Genome Res.* **9**, 27–43 (1999).
- Feng, H. P. & Gierasch, L. M. Molecular chaperones: clamps for the Clps? *Curr. Biol.* **8**, 464–467 (1998).
- Traut, T. W. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide binding sites. *Eur. J. Biochem.* **222**, 9–19 (1994).
- Saraste, M., Sibbald, P. R. & Wittinghofer, A. The P-loop—a common motif in ATP—and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430–434 (1990).
- Smith, C. A. & Rayment, I. Active site comparisons highlight structural similarities between myosin and other P-loop proteins. *Biophys. J.* **70**, 1590–1602 (1996).
- Rohrwild, M. *et al.* The ATP-dependent HslVU protease from *Escherichia coli* is a four-ring structure resembling the proteasome. *Nature Struct. Biol.* **4**, 133–139 (1997).

19. Schirmer, E. C., Glover, J. R., Singer, M. A. & Lindquist, S. Hsp100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem. Sci.* **21**, 289–296 (1996).
20. Karata, K., Inagawa, T., Wilkinson, A. J., Tatsuta, T. & Ogura, T. Dissecting the role of a conserved motif (the second region of homology) in the AAA family of ATPases. *J. Biol. Chem.* **274**, 26225–26232 (1999).
21. Levchenko, I., Smith, C. K., Walsh, N. P., Sauer, R. T. & Baker, T. A. PDZ-like domains mediate binding specificity in the Clp/Hsp100 family of chaperone. *Cell* **91**, 939–947 (1997).
22. Smith, C. K., Baker, T. A. & Sauer, R. T. Lon and Clp family proteases and chaperones share homologous substrate-recognition domains. *Proc. Natl Acad. Sci.* **96**, 6678–6682 (1999).
23. Weber-Ban, E. U., Reid, B. G., Miranker, A. D. & Horwich, A. L. Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *Nature* **401**, 90–93 (1999).
24. Löwe, J. *et al.* Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **268**, 533–539 (1995).
25. Groll, M. *et al.* Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **386**, 463–471 (1997).
26. Knowlton, J. R. *et al.* Structure of the proteasome activator REG α (PA28 α). *Nature* **390**, 639–643 (1997).
27. Glickman, M. H. *et al.* A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* **94**, 615–623 (1998).
28. Project, C. C. C. The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760–763 (1994).
29. Jones, A. T. A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272 (1978).
30. Brünger, A. T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Cryst. D* **54**, 905–921 (1998).

Acknowledgements

We thank P. Zwickl for ^{14}C -labelled casein, for his help with the protein degradation assay and for essential discussions; S. Grazulis for sharing PDB-scripts M. Boicu for DNA sequencing; and R. Engh for critically reading the manuscript. The financial support of the Deutsche Forschungsgemeinschaft and of the Human Frontier Science Program is acknowledged.

Correspondence and requests for materials should be addressed to M. B. (bochtler@biochem.mpg.de). Coordinates have been deposited in Protein Data Bank under accession numbers 1DOO (P6322), 1DO0 (P21212) and 1DO2 (P321).

Structure of the intact transactivation domain of the human papillomavirus E2 protein

Alfred A. Antson^{*†}, Julie E. Burns^{†‡}, Olga V. Moroz^{*}, David J. Scott^{*§}, Cyril M. Sanders^{‡||}, Igor B. Bronstein^{*}, G. Guy Dodson^{*¶}, Keith S. Wilson^{*} & Norman J. Maitland[‡]

^{*}York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK

[‡]YCR Cancer Research Unit, Department of Biology, University of York, York YO10 5DD, UK

[§]Department of Biology, University of York, York YO10 5DD, UK

[¶]National Institute for Medical Research, London NW7 1AA, UK

[†]These authors contributed equally to this work

Papillomaviruses cause warts and proliferative lesions in skin and other epithelia. In a minority of papillomavirus types ('high risk', including human papillomaviruses 16, 18, 31, 33, 45 and 56), further transformation of the wart lesions can produce tumours¹. The papillomavirus E2 protein controls primary transcription and replication of the viral genome². Both activities are governed by a ~200 amino-acid amino-terminal module (E2NT) which is connected to a DNA-binding carboxy-terminal module by a flexible linker. Here we describe the crystal structure of the complete E2NT module from human papillomavirus 16. The

E2NT module forms a dimer both in the crystal and in solution. Amino acids that are necessary for transactivation are located at the dimer interface, indicating that the dimer structure may be important in the interactions of E2NT with viral and cellular transcription factors. We propose that dimer formation may contribute to the stabilization of DNA loops³ which may serve to relocate distal DNA-binding transcription factors to the site of human papillomavirus transcription initiation.

Human papillomaviruses (HPVs) have evolved a sophisticated system of control, mediated by protein–DNA and protein–protein interactions, that involves both cellular and viral proteins. The relative molecular mass 45,000 (M_r 45K) nuclear phosphoprotein, E2, has two key roles in this control. It functions as the principal virally encoded transcription factor and, in association with the viral E1 protein, it creates the molecular complex at the origin of viral DNA replication⁴. The E2 protein has three distinct modules (Fig. 1). The sequences of the N-terminal module (E2NT) of ~200 amino acids are highly conserved among the 86 known papillomavirus types (E2 Sequence Database, <http://hpv-web.lanl.gov>; see Supplementary Material), and this module is responsible for interactions with viral and host-cell transcription factors, such as Sp1, TFIIB and AMF-1 (refs 2, 5, 6). It is followed by a flexible, proline-rich, linker module and a C-terminal module (E2CT), each of ~90 amino acids. E2CT binds tightly as a homodimer to DNA sites with a consensus sequence of ACCGN₄CGGT (ref. 2). The crystal structure of E2CT determined in complex with cognate DNA⁷ shows that the protein dimer induces substantial bending (42–44°) of the DNA from its B-form double helix. The structure of a proteolytic fragment of HPV18 E2NT has been determined at 2.1 Å resolution⁸.

Here we report the crystal structure of the complete HPV16 E2NT module refined with data extending to 1.9 Å spacing (Table 1). There are two domains, N1 and N2, arranged to give the module an overall L-shaped appearance. The N1 domain (residues 1–92) is composed of three long α -helices (Fig. 2), packed antiparallel to one another in the form of a twisted plane, with angles of about 20° and 25° between consecutive helices. There is a tight loop between helices α 1 and α 2 and a more extended one between helices α 2 and α 3. There is no structurally homologous domain in the PDB⁹. Domain N2 comprises residues 110–201 and is almost entirely antiparallel β -sheet structure. Our results and those reported for the HPV18 fragment⁸ show that the N2 domain also has a new fold. Between the N1 and N2 domains, residues 93–109 make up two consecutive single turns of helical structure forming a fulcrum which packs tightly against both domains and could be considered to be part of either N1 or N2. Thirty-three of the total 201 residues in our E2NT construct are totally or highly conserved. Twenty-five of these appear to have a purely structural role, with 12 clustering at the N1–N2 interface reflecting the functional importance of the fulcrum in defining the relative orientation of the two domains (Fig. 2). Eight conserved amino acids are exposed to the solvent, indicating that they may have a role in intermolecular interactions. Mutational substitutions (reviewed in refs 2, 10) indicate that two of these residues are involved in transactivation (Arg 37 and Ile 73) and one in replication (Glu 39).

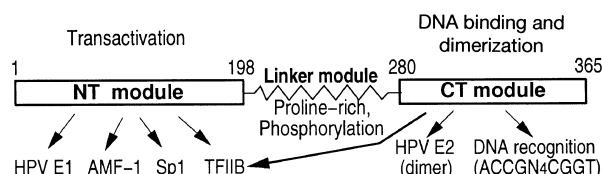


Figure 1 Functional and structural assignments of papillomavirus E2. View of the N-terminal, linker and C-terminal modules indicating the known functions of each. Amino-acid numbers that delimit the modules correspond to those of HPV16 E2.

^{||}Present address: Cold Spring Harbor Laboratory, PO Box 100, Bungtown Road, Cold Spring Harbor, New York 11724, USA.

The structure of the proteolytic fragment of HPV18 E2NT, which has 65 N-terminal residues (two-thirds of the N1 domain) missing, has allowed some analysis of mutational effects on function⁸. The fragment, however, lacks several key residues that are essential for the transcriptional and replication activities and that are, see below, essential for E2NT dimer formation. Human papillomavirus 18 E2NT has an overall sequence identity of 47% with HPV16; superposition of their N2 domains shows an r.m.s. difference of less than 1.0 Å and a relative rotation of the N1 and N2 domains of ~8° (Fig. 2). This may suggest that there is some difference or flexibility in their relative orientation about the fulcrum, but it may also result from the missing part of N1 in the HPV18 fragment. The C-terminal residues of the two structures, from residue 196 onwards, have quite distinct conformations with the chain pointing in opposite directions (Fig. 2).

A striking feature of the crystal structure is the association of the intact HPV16 E2NT into a dimer, with two monomers packed around the crystallographic twofold axis (Fig. 3a, b). The dimer interface is formed mostly by amino acids from helices α2 and α3 of the N1 domain and by residues 142–144 from N2. The total buried surface area between the two E2NT modules is 2,026 Å², comparable to the 2,444 Å² buried between the two E2CT modules⁷, which are known to form a tight dimer with a dissociation constant (K_d) of 3–6 × 10⁻⁸ M^{11,12}. In the E2NT dimer interface, each subunit contributes a cluster of seven residues (Arg 37, Ala 69, Ile 73, Gln 76, Leu 77, Thr 81 and Glu 80), which are invariant or conserved, with many direct or water-mediated hydrogen bonds and rather few non-polar contacts (Fig. 3a). Indeed, all the direct hydrogen bonds

between monomers are made through these seven amino acids. The conservation or conservative substitution of residues at the monomer–monomer interface in the known sequences suggests that the E2NT dimer interactions are present in all known papillomaviruses. To verify whether the crystallographically observed dimer forms in solution, we measured the dissociation constant by sedimentation equilibrium using analytical ultracentrifugation (see Supplementary Material). The resulting value of K_d , $8.1 \pm 4.2 \times 10^{-6}$ M, indicates medium-strength association. The micromolar range of K_d is physiologically significant, and comparable to values for other transcription factors which often have K_d values between 1 μM and 20 μM^{13,14}.

In HPV, the genes are preceded by a long upstream regulatory region (URR) that usually contains four spatially separate E2-binding sites (Fig. 4a). The fundamental and very stable interaction of E2 with the viral URR sites is through the E2CT dimer, with K_d often in the nanomolar range¹⁵. Electron microscopy studies have shown that intact E2 led to the formation of loops on DNA templates with the E2-binding sites up to 560-base pairs apart, whereas a truncated E2, containing the DNA-binding E2CT but missing the N-terminal 161 residues, did not³. It was deduced that E2NT modules from distantly bound E2 dimers interact with each other sufficiently strongly to stabilize the DNA loops. The existence of such interdimer interactions was further supported by the observation of synergistic transcription activation by a complex of two DNA-bound E2 dimers¹⁶. Our data suggest that E2NT mediates these interactions via the formation of E2NT 'interdimers' (Fig. 4b) with the same molecular contacts between E2NT modules as seen in

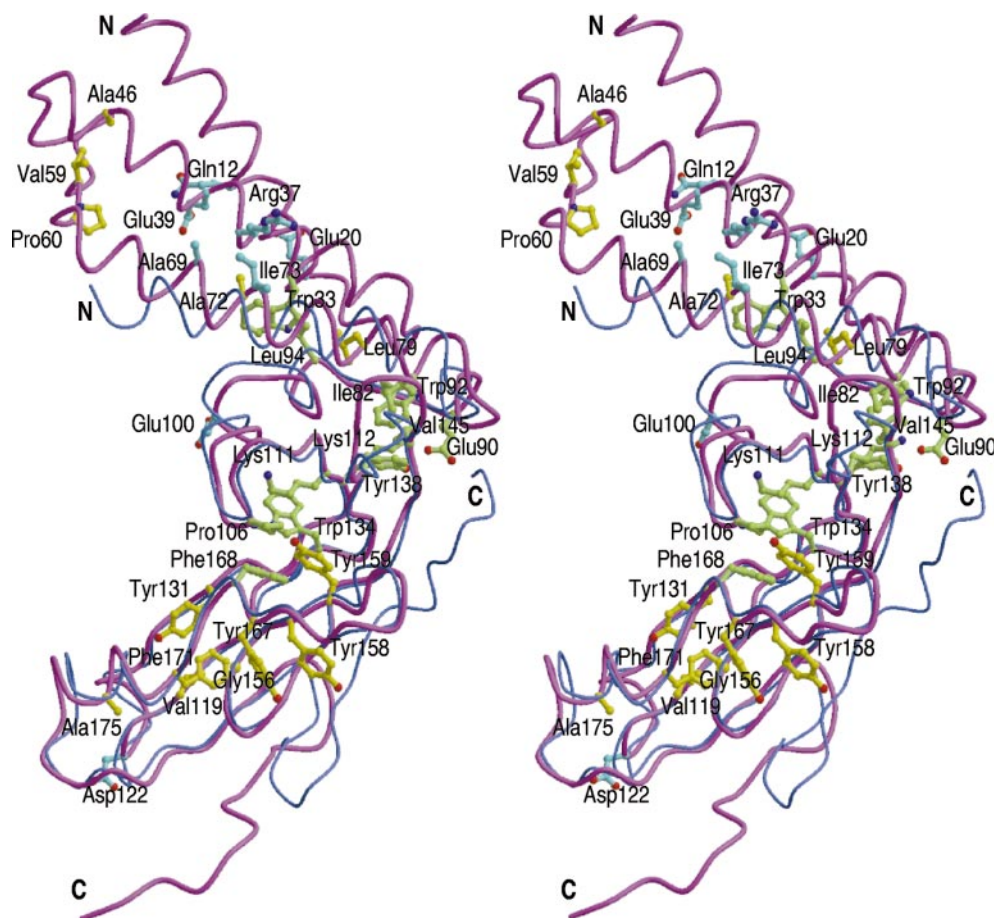


Figure 2 Stereo diagram showing the distribution of conserved residues within the E2NT monomer. Side chains of internal structural residues in the N1 and N2 domains are shown in yellow, and those in the interface between the two domains are shown in green. Side

chains of residues exposed on the surface with a potential for protein–protein interactions are shown in aquamarine. The structure is overlaid with that of the backbone HPV18 E2NT fragment⁸ (blue). Generated with BOBSCRIPT^{27,28} and Raster3D²⁹.

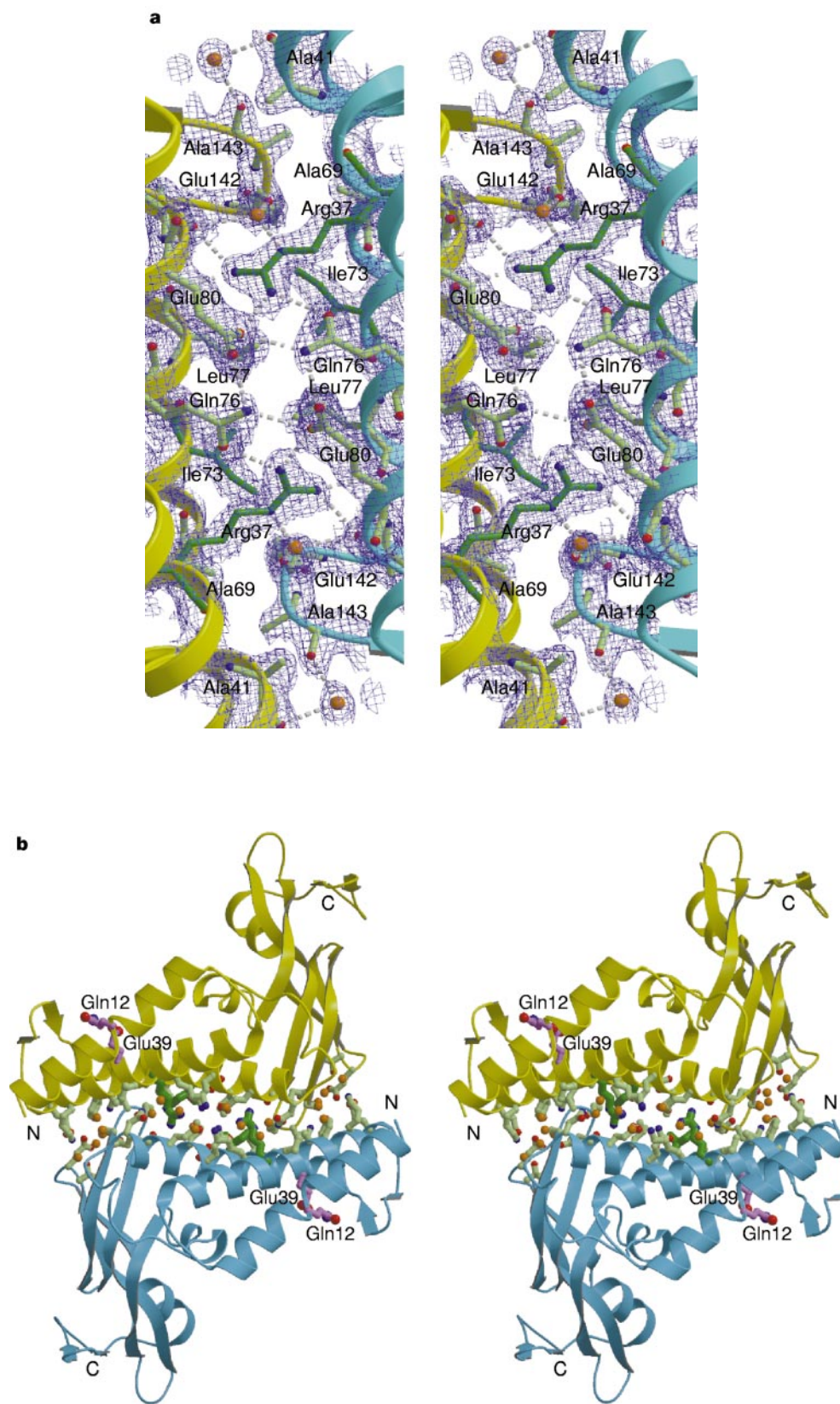


Figure 3 Structural features of E2. **a**, Stereo view of the electron density with the final model, at the dimer interface of the E2NT module, viewed down the crystallographic twofold axis. The likelihood weighted $2|F_o| - |F_c|$ electron density map is contoured at the 1.5σ level. Ribbons of two independent monomers are shown in aquamarine and yellow. Side chains of Arg 37 and Ile 73, which are known to be critical for transactivation^{2,10}, are shown in dark green; side chains of other residues at the dimer interface are shown in light

green. Oxygen atoms are in red, nitrogen atoms in blue; water molecules are shown as orange spheres and hydrogen bonds as dashed sticks. **b**, Stereo ribbon diagram of the E2NT dimer, showing the extent of the interface between the two subunits. The view is as in **a** but rotated clockwise by 90° . Side chains of Gln 12 and Glu 39 that are critical for interactions with E1 (refs 2,10,18) are shown in magenta. Side chains of residues at the dimer interface are coloured as in **a**.

the crystal structure. Functional importance of the E2NT dimer is further supported by mutation of Arg 37, Gln 76 or Ile 73 (all at the dimer interface) which resulted in considerable disruption of transactivation, but had little effect on replication^{2,8,10}. This introduces the hypothesis that it may be disruption of the dimer interface that affects transactivation rather than direct interaction of these residues with associated factors. The two E2NT modules in one E2 molecule could potentially bind either to one another ('intradimer') or to E2NT modules from another E2 dimer bound to a separate DNA site ('interdimer'). It is not clear whether E2NT modules in an intact E2 dimer form an intradimer interaction before the development of DNA loops; the existence of an intradimer would require domain exchange to take place during loop formation.

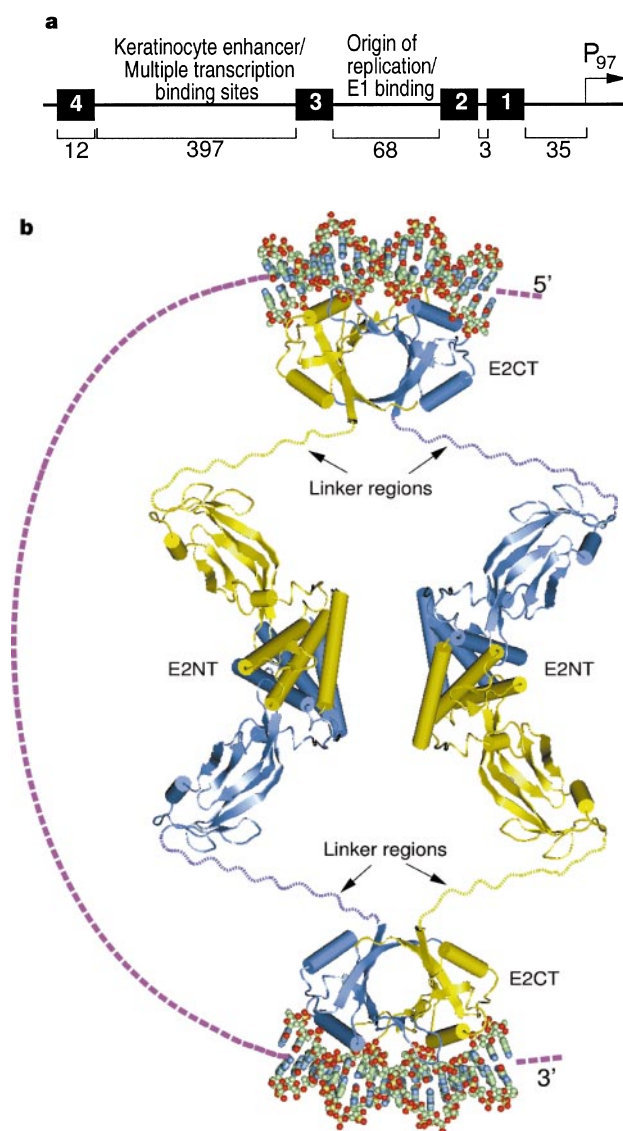


Figure 4 Loop formation in the upstream regulatory region (URR) of HPVs. **a**, Schematics of the URR of the HPV16 genome containing the four E2-binding sites (black boxes 1–4); all four sites are 12 bases long. The numbers at the bottom of the scheme indicate the length of the elements of the URR upstream of the p97 promoter. **b**, A model of the formation of DNA loops, showing interactions between both the E2NT and E2CT modules. The E2NT dimer is from the present study. The E2CT dimer, with its bound DNA, is based on the crystal structure of this module⁷. The relative orientation and position of the E2NT and E2CT dimers is purely schematic. The two subunits constituting the E2NT and E2CT dimers are shown as yellow and blue ribbons; DNA is shown in ball-and-stick representation. Generated using QUANTA (Molecular Simulations).

Replication of the viral genome is initiated by the binding of another viral protein, E1, to the origin of DNA replication², which is itself flanked by two E2-binding sites (Fig. 4a). Whereas the function of E2CT dimers is to bind specifically to the DNA sites, E2NT interaction with E1 enhances the binding of E1 to this region. Mutational substitutions of the conserved Glu 39 substantially reduced DNA replication, but transcriptional activation was generally retained^{2,10}. In the structure, Glu 39 makes every possible hydrogen bond through its side-chain carboxyl oxygens, three to water molecules and one to NE2 of the conserved Gln 12. Although mutation of Gln 12 in bovine papillomavirus type 1 (BPV1) only slightly affected both transactivation and replication, it substantially reduced cooperative origin binding^{17,18}. The close proximity of Gln 12 and Glu 39 further strengthens the notion that these two residues are involved in interactions with E1 and hence replication. The Gln 12/Glu 39 cluster lies on the opposite side of the N1 domain to that involved in transactivation and dimerization (Fig. 3b). The association sites of E2NT with cellular transcription factors AMF-1 (residues 74–134) and TFIIB (134–216) are distinct^{5,6}. In the structure, residues 74–134 include the fulcrum, whereas residues 134–216 correspond to domain N2.

In conclusion, our results indicate that dimerization of E2NT modules may link E2 dimers and may serve as the key to the formation of DNA loops (Fig. 4b), the mechanism by which distally bound transcription factors would be brought close to the site of transcription initiation. In addition, our results raise the possibility that dimer formation may serve as a molecular switch between early gene expression and viral genome replication during HPV infection¹⁹. One scenario would be to activate transcription through the induction of DNA loop formation at early stages of the viral life cycle. At later stages, when the concentration of expressed E2 protein within the cell becomes high and comparable to the K_d for E2NT dimer formation, free E2NT modules could compete for dimerization with those involved in DNA loop formation, switching off transcription and stimulating replication; other protein factors could be involved in this process, including E1. Understanding the complex structural events involved in transactivation and replication of papillomaviruses requires further studies of E2 interactions

Table 1 X-ray structure determination

X-ray data and phasing Data	Native	Uranyl acetate	Gold cyanide
Resolution Å	30.0–1.9 (1.93–1.90)	20.0–2.7 (2.74–2.7)	20.0–2.7 (2.74–2.7)
(outer shell)			
Unique reflections	21,751	7,873	7,937
Redundancy*	3.7	7.0	6.8
Completeness (%)	98.8 (89.3)	99.9 (99.8)	100.0 (100.0)
$R_{\text{merge}}^{\dagger}$	0.058 (0.339)	0.073 (0.271)	0.061 (0.268)
Phasing power [‡] (centric/acentric)	–	1.55/2.07	0.95/1.40
Refinement and model correlation			
Resolution range			1.9–30.0 Å
Number of protein atoms			1,622
Number of solvent sites			211
Number of reflections used in refinement			20,637
Number of reflections used for R_{free} calculation			1,111
R -factor [§]			0.232
$R_{\text{free}}^{\S}$			0.305
Average atomic B -factor (Å ²)			
Protein atoms			38.0
Water molecules			48.5
R.m.s. deviations from ideal geometry (targets in parentheses)			
Bond distance (Å)			0.013 (0.020)
Angle distance (Å)			0.026 (0.040)

* The average number of observations of the same reflection.

[†] The value of the merging R -factor between equivalent measurements of the same reflection, $R_i = \sum |I_i - \langle I \rangle| / \sum I_i$.

[‡] Phasing power, $P = \sum |F_h| / \sum |F_p| \pm F_p - F_h$.

[§] Crystallographic R -factor ($R_{\text{free}} = \sum |F_o| - |F_c| / \sum |F_o|$).

[§] I_i , Observed intensity of a reflection; amplitudes of structure factors are designated as F_h for heavy atom, F_p for protein plus heavy atom, F_o for observed and F_c for calculated.

with viral and cellular transcription factors. The dimerization surface of E2 could prove to be a good target for designing drugs against papillomaviruses because E2 is essential for viral transcription, there is no homologous human protein and the residues forming the E2NT dimer interface are highly conserved among different viral strains. □

Methods

Purification and crystallization

Details of the purification and crystallization of E2NT have been described²⁰. The best crystals were obtained by hanging-drop vapour diffusion with protein at 4.0–5.0 mg ml⁻¹ in 10 mM Tris-HCl pH 8.0, 5 mM DTT, 0.2 mM EDTA, 300 mM NaCl. The reservoir contained 11% monomethylether PEG 5000, 5–9% isopropanol, 0.1 M triethanolamine pH 8.3 and 300 mM NaCl. Crystals (space group *P*₃21 with *a* = *b* = 54.7 Å, *c* = 155.7 Å) grew only with very fresh protein preparations and deteriorated in terms of diffraction quality in less than a week, which necessitated vitrifying and storing all crystals in liquid nitrogen immediately after growth. The cryo-solution contained 15% PEG 5000 monomethylether, 15% 2-methyl-2,4-pentanediol, 20 mM Tris-HCl pH 8.0, 0.1 M triethanolamine pH 8.3 and 300 mM NaCl.

Structure determination

Data on the native protein and two derivatives were measured at EMBL Hamburg beam line X11 at a wavelength of 0.89 Å and merged using SCALEPACK²¹ (Table 1). The two derivatives, prepared using 2 mM uranum acetate and 2 mM gold cyanide, were solved using the CCP4 suite²² from the difference Patterson synthesis and by direct methods as implemented in SHELX²³. Phases were calculated at 2.7 Å resolution using MLPHARE, with a figure of merit of 0.59, and extended to 1.9 Å using DM²⁴. An initial model was built using QUANTA (Molecular Simulations) and XFIT²⁵. The model was completed with REFMAC (resolution 30–1.9 Å), to an *R*-factor of 23.2% (*R*_{free} 30.5%) (Table 1). There are 221 residues in the recombinant protein, the first twenty comprise the 6×His tag. The refined model contains all but 2 of the 201 residues of the real protein: residues 125–126 are disordered in a flexible surface loop. Only one residue (His0) of the 6×His tag has clear density. All amino acids lie in the allowed region of the Ramachandran plot with 92.4% in the most favoured regions.

Analytical ultracentrifugation

Experiments were carried out in an Optima XL-A ultracentrifuge (Beckman-Coulter) using scanning ultraviolet optics (see Supplementary Material). Three concentrations of recombinant E2NT were loaded into cells containing 12 mm pathlength six-channel Epon centrepieces with quartz windows. The solvent was 10 mM Tris-HCl pH 8.0, 5 mM dithiothreitol, 0.2 mM EDTA and 300 mM NaCl. The solvent density and partial specific volume of the protein were calculated as 1.01128 and 0.7284, respectively, using the program SEDNTERP, according to published methods²⁶. Data were obtained at rotor speeds of 12,000 and 16,000 r.p.m., and the time to equilibrium was typically 10–12 h. All runs were carried out at 293 K, and all radial scans were at a wavelength of 280 nm. The dissociation constant was obtained by nonlinear regression using the Beckman ultracentrifuge software, and the variance of the fit was found to be 1.98 with 988 degrees of freedom.

Received 22 July; accepted 8 December 1999.

- zur Hausen, H. Molecular pathogenesis of cancer of the cervix and its causation by specific human papillomavirus types. *Curr. Top. Microbiol. Immunol.* **186**, 131–156 (1994).
- McBride, A. & Myers, G. in *Human Papillomaviruses 1997* (eds Myers, G. et al.), III–54–III–73 (Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico, 1997).
- Knight, J. D., Li, R. & Botchan, M. The activation domain of the bovine papillomavirus E2 protein mediates association of DNA-bound dimers to form DNA loops. *Proc. Natl Acad. Sci. USA* **88**, 3204–3208 (1991).
- Mohr, I. J. et al. Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. *Science* **250**, 1694–1699 (1990).
- Breiding, D. E. et al. Functional interaction of a novel cellular protein with the papillomavirus E2 transactivation domain. *Mol. Cell. Biol.* **17**, 7208–7219 (1997).

- Yao, J. M., Breiding, D. E. & Androphy, E. J. Functional interaction of the bovine papillomavirus E2 transactivation domain with TFIIB. *J. Virol.* **72**, 1013–1019 (1998).
- Hegde, R. S., Grossman, S. R., Laimins, L. A. & Sigler, P. B. Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* **359**, 505–512 (1992).
- Harris, S. F. & Botchan, M. R. Crystal structure of the human papillomavirus type 18 E2 activation domain. *Science* **284**, 1673–1677 (1999).
- Bernstein, F. C. et al. The protein data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542 (1977).
- Cooper, C. S., Upmeyer, S. N. & Winokur, P. L. Identification of single amino acids in the human papillomavirus 11 E2 protein critical for the transactivation or replication functions. *Virology* **241**, 312–322 (1998).
- Mok, Y. K., Gay, G. D., Butler, P. J. & Bycroft, M. Equilibrium dissociation and unfolding of the dimeric human papillomavirus strain-16 E2 DNA-binding domain. *Protein Sci.* **5**, 310–319 (1996).
- Foguel, D., Silva, J. L. & de Prat-Gay, G. Characterization of a partially folded monomer of the DNA-binding domain of human papillomavirus E2 protein obtained at high pressure. *J. Biol. Chem.* **273**, 9050–9057 (1998).
- Estojak, J., Brent, R. & Golemis, E. Correlation of two-hybrid affinity with *in vitro* measurements. *Mol. Cell. Biol.* **15**, 5820–5829 (1995).
- Sengchanthalangsy, L. et al. Characterisation of the dimer interface of transcription factor NKκB p50 homodimer. *J. Mol. Biol.* **289**, 1029–1040 (1999).
- Chao, S.-F. et al. Subunit affinities and stoichiometries of the human papillomavirus type 11 E1:E2:DNA complex. *Biochemistry* **38**, 4586–4594 (1999).
- Gauthier, J.-M., Dostatni, N., Lusky, M. & Yaniv, M. Two DNA-bound E2 dimers are required for strong transcriptional activation and for cooperation with cellular factors in most cells. *The New Biologist* **3**, 498–509 (1991).
- Abroi, A., Kurg, R. & Ustav, M. Transcriptional and replicational activation functions in the bovine papillomavirus type 1 E2 protein are encoded by different structural determinants. *J. Virol.* **70**, 6169–6179 (1996).
- Brokaw, J. L., Blanco, M. & McBride, A. A. Amino acids critical for the functions of the bovine papillomavirus type 1 E2 transactivator. *J. Virol.* **70**, 23–29 (1996).
- Desaintes, C. & Demeret, C. Control of papillomavirus DNA replication and transcription. *Semin. Cancer Biol.* **7**, 339–347 (1996).
- Burns, J. E. et al. Expression, crystallization and preliminary X-ray analysis of the E2 transactivation domain from papillomavirus type 16. *Acta Crystallogr. D* **54**, 1471–1474 (1998).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Collaborative Computational Project N4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Sheldrick, G. M. & Schneider, T. R. SHELXL: high-resolution refinement. *Methods Enzymol.* **277**, 319–343 (1997).
- Cowan, K. D. & Main, P. Phase combination and cross-validation in iterated density-modification calculations. *Acta Crystallogr. D* **52**, 43–48 (1996).
- Oldfield, T. J. in *Proceedings of the CCP4 Study Weekend* (eds Bailey, S., Hubbard, R. & Waller D.) 15–18 (Daresbury Laboratory, Warrington, UK, 1994).
- Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (eds Harding, S. E., Rowe, A. J. & Horton, J. C.) 90–125 (Royal Society of Chemistry, London, 1992).
- Kraulis, P. J. MOLSCRIPT—a program to produce both detailed and schematic plots of proteins structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Esnouf, R. M. An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics* **15**, 133–138 (1997).
- Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial offices of Nature.

Acknowledgements

K.S.W., G.G.D., A.A.A. and O.V.M. thank the BBSRC for infrastructure support and the EC for supporting work at EMBL, Hamburg, through the Access to Large Installations Project, Contract Number CHGE-CT93-0040. We also thank Yorkshire Cancer Research for project and program grant support to N.J.M., J.E.B., I.B.B. and O.V.M. A.A.A. is supported by a Wellcome Trust Career Development Fellowship.

Correspondence and requests for materials should be addressed to K.S.W. (e-mail: keith@yorvic.york.ac.uk). The structure factors and the refined coordinates of the E2NT module have been deposited in the Protein Data Bank⁹ under accession codes 1R1DTOSF and 1DITO, respectively.

Biochips on a plate

Microarrays in their many guises are the main focus this week.

CMT-GAPS

Corning Science Products

www.corning.com

Coated slide for printing microarrays

Measuring 25 mm × 75 mm and made of glass, these slides come with an aminosilane coating designed to ensure the even immobilization of DNA for printed microarrays. Uniform spot morphology, high DNA retention for maximum signal strength, and ultra low fluorescent background are claimed for CMT-GAPS coated slides. Every slide is certified RNase and DNase free and each one is tested for consistent contact angle, visually inspected and subjected to a rigorous cleaning process.

Reader Service No. 100

Strip-EZ RT

Ambion

www.ambion.com

Strip and reprobe gene arrays at least 10 times without significant loss of target

In the Strip-EZ RT kit, modified nucleotides are incorporated into cDNA probes along with the desired label. After hybridization a reagent specifically degrades the modified nucleotides. The resulting probe fragments are gently removed from the gene arrays without damage to the membrane-bound nucleic acids. The kit contains reagents for probe synthesis and removal.

Reader Service No. 101

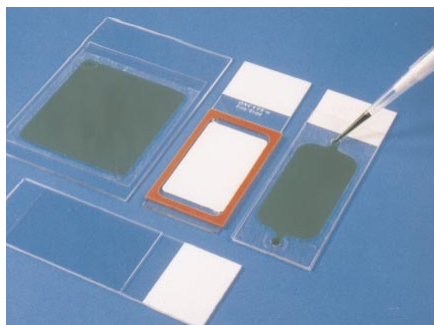
Hybridization chambers

Grace Bio-Labs

www.gracebio.com

For high-throughput screening

The new line of hybridization chambers from Grace Bio-Labs includes a range of sizes and styles for high throughput screening of genomic arrays on microscope slides, glass plates, Film-Slides, and blotting membranes. Four styles designed to minimize reagent volume, eliminate evaporation and



Grace notes: a choice of chambers.



Next slide please: Corning's glass for microarrays.

promote reagent mixing and uniform hybridization. The nuclease-free, ready-to-use chambers are temperature resistant and compatible with commonly used hybridization reagents. The chambers are suitable for protocols for autoradiographic, fluorescent or chemiluminescent endpoints.

Reader Service No. 102

OpArrays

Operon Technologies

www.operon.com

Optimized, customizable microarrays

Operon's synthesis and bioinformatics groups have jointly developed OpArrays, based on proprietary technologies. The resulting highly optimized, customizable arrays feature covalent, stable chemistry, customized layout, and are sequence optimized to minimize cross-hybridization. Operon offer technical support to help customers design their own arrays. Medium-density microarrays may be constructed with optimized single-stranded oligonucleotides or double-stranded PCR products. Operon have also introduced a new line of stock arrays, including human and yeast collages to cover commonly studied genes such as those involved with apoptosis, transcription factors, signal transduction, and DNA repair.

Reader Service No. 103

Atlas arrays

Clontech

www.clontech.co.uk

Human and human cancer 1.2 arrays

New from Clontech, the Atlas Human and Atlas Human Cancer 1.2 Arrays are expanded membrane-based arrays for gene expression profiling. Each array contains 1,176 unique cDNA fragments. Gene classes included in previous versions are more extensively covered and several new gene categories have been added. Atlas Arrays consist of unique cDNA fragments of well-characterized genes arrayed on nylon mem-

branes. The new 1.2 kits supply four membranes of single-spotted genes, allowing simultaneous comparison of gene expression profiles of two different samples in a single experiment.

Reader Service No. 104

Lumi Light^{PLUS}

Roche Diagnostics

www.roche.com

A new peroxidase (POD) chemiluminescence substrate

Lumi Light^{PLUS} Western Blotting Substrate is claimed to be more sensitive and longer-lasting than any other POD substrate on the market. Lumi Light allows for the detection of very low amounts of antigen blotted to nitrocellulose or PVDF membrane using primary antigen specific antibodies, and either secondary anti-mouse or anti-rabbit IgG antibodies, conjugated with peroxidase. The Lumi Light substrate remains stable for more than nine hours without any significant loss in intensity. Savings on primary antibody use can be made, as it is possible to dilute Lumi Light 10–100-fold more than with calorimetric detection systems thanks to Lumi Light's strong signal. The final detection process can produce multiple exposures, since exposures of only a few minutes are necessary. Quantification of the Lumi Light signal can be simply carried out at a wavelength of 425 nm using either an X-Ray film or an imager.

Reader Service No. 105

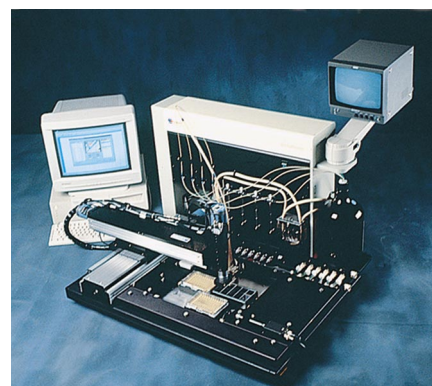
BioChip Arayer

Packard Instrument Company

www.packardinst.com

New options for this product line

The BioChip Arayer, an automated pipetting platform for the development of microarrays, now includes the option of cooling the microplate and slide positions



Manifold benefit: robotic handling from Packard.

new on the market

through the use of a recirculating fluid. A manifold distributes user-supplied cooled fluid to specialized microplate and slide adapters. In addition, a vacuum manifold is provided which delivers vacuum pull to each position, ensuring there is no movement of the microplates or slides during robotic processing. These features, combined with the ability of the PiezoTip to deliver precise nanolitre volumes at micrometre spacing, help minimize evaporation of sample and maintain discrete orthogonal arrays on porous and non-porous substrates.

Reader Service No. 106

XNA on Gold

Interactiva

www.interactiva.de

Suitable for a wealth of applications in genomics, proteomics or glycomics

The modular biochip XNA on Gold has been commercially launched through a Technology Access Program by Interactiva Biotechnologie GmbH. The new affinity array biochip is designed to ask many molecular questions out of a single sample. Interactiva claim the biochip is extremely versatile, thanks to robust immobilization, automated processes, good sensitivity, easy detection and good usability with either genes or proteins. Sensor elements ranging from nucleic acids to proteins could be applied to Interactiva's chips thanks to universal streptavidin-biotin binding.

Reader Service No. 107

OmniGrid

GeneMachines

www.genemachines.com

A new option for the OmniGrid microarrayer

This new option, a hard-sided enclosure, includes a HEPA filter to reduce dust levels during microarray operation, eliminating the problems caused by the intervention of

dust. This new option also allows the user to control humidity levels to prevent evaporation during spotting. The OmniGrid slide platter is accessible through one hinged door and two sliding side doors in the enclosure, allowing easy access to sample plates and slides.

Reader Service No. 108

Scanning service

NEN

www.nenlifesci.com

High-volume differential gene-expression analysis without special equipment

Slide scanning and data processing services are now available from NEN Life Science Products for Micromax cDNA microarray systems. The service enables researchers to use the glass slide-based microarray systems for high throughput, differential gene expression analysis, without investing in an expensive laser scanning system. NEN uses a ScanArray confocal laser scanner (from GSI Lumonics Inc.) to scan customers' slides. Cyanine-3 and cyanine-5 channels are scanned sequentially, and results are saved to both JPG and 16-bit TIFF files. In order to maximize the dynamic range of reliable expression data, scans are performed at both low/medium and high laser power and photomultiplier settings, yielding two complete sets of data.

Reader Service No. 109

GeneTAC 2000

Genomic Solutions

www.genomicsolutions.com

Making analysis of microarrays easier

The GeneTAC 2000 BioChip Analyzer utilizes a powerful detector to capture images with a 16-bit dynamic range. The sample carousel automates imaging and analysis of up to 24 fluorescently labelled biochips at one time. Quantum efficiency of the detector



Image conscious: the GeneTAC 2000 analyser.

is 60–70%. Genomic Solutions claim that array analysis which used to take several hours can be done in less than 30 minutes. The software automates spot detection and GeneTAC 2000 is fitted with a fully integrated analysis package which uses a powerful database (Microsoft SQL Server 7.0) for bioinformatics applications.

Reader Service No. 110

3D-Link

SurModics

www.surmodics.com

Activated slides for DNA microarrays

3D-Link Activated Slides are designed to simplify the production of high-quality microarrays. Features listed include a uniform consistent coating to prevent slide and lot variability, low backgrounds, stable and covalent bond between the DNA and the surface of the slide, and endpoint attachment for the captured DNA.

Reader Service No. 111

BioChipReader

LaVision

www.lavision.com

Rapid readout for biochips

LaVision's sensitive (≥ 0.1 fluorophore per μm^2) and fast-readout system for biochips incorporates an excitation source and optics to allow simultaneous excitation and detection at different wavelengths. The software handles device control, chip read-out, data processing and analysis. The BioChipReader can be customized both in terms of hardware and software. Exposure time is 1 ms–1,000 s, spatial resolution better than 10,000 dots per chip, chip size either $3 \times 3 \text{ mm}^2$ or $80 \times 120 \text{ mm}^2$, and emission 400–900 nm with multi-colour detection. Applications for the chip reader include drug screening, sequencing by hybridization, cell screening and epitope mapping. The LaVision website was "under construction" as *Nature* went to press.

Reader Service No. 112

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card.

ADVERTISEMENTS

**BACs
YACs
PACs
cDNAs**

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

Check out our homepage at <http://www.resgen.com>

**Biomolecular Synthesis
oligos, synthetic
peptides and polyclonal
antibodies**

Contact Severn Biotech Ltd
Tel: 01562 825 286
Fax: 01562 825 284
email: info@severnbiotech.com
www.severnbiotech.com

READER ENQUIRY NO. 6

READER ENQUIRY NO. 28